



Assessing the importance of the choice threshold in quantifying market risk under the POT approach (EVT)

Sonia Benito¹ · Carmen López-Martín¹ · M^a Ángeles Navarro¹

Accepted: 1 September 2022 / Published online: 6 January 2023
© The Author(s), under exclusive licence to Springer Nature Limited 2023

Abstract

From a theoretical point of view, the selection of thresholds is a critical issue in the framework of the Peaks Over Threshold (POT) approach, which is why in the last decade numerous methodologies have been proposed for its selection. In this paper, we address this subject from an empirical point of view by assessing to what extent the selection of the threshold is decisive in quantifying the market risk. For measuring market risk, we use the Value at Risk (VaR) and the Expected Shortfall (ES) measures. The results obtained indicate that there is a large set of thresholds that provide similar Generalized Pareto Distribution (GPD) quantiles estimators and as a consequence similar market risk measures. Just only, for large thresholds, those corresponding to the 98th and 99th percentile of the GPD some differences are found. It means that the choice of threshold in the framework of the POT method may not be relevant in quantifying market risk when we use the VaR and ES measures for this task.

Keywords Extreme value theory · Peaks over threshold · Value at risk · Expected shortfall · Generalized Pareto distribution

Introduction

Extreme value analysis has wide applications in many fields, such as civil engineering (Wu and Qiu 2018), climatology (Davison and Smith 1990; Kharin et al. 2013), seismology (Beirlant et al. 2018), hydrology (Katz et al. 2002; Carreau et al. 2017;

✉ Sonia Benito
soniabm@cee.uned.es

Carmen López-Martín
carmen.lopez@cee.uned.es

M^a Ángeles Navarro
mnavarro900@alumno.uned.es

¹ Faculty of Economics and Business Administration, National Distance Education University (UNED), Senda del Rey, 11, 28040 Madrid, Spain



Bader et al. 2018), insurance (Reiss and Thomas 2007) and finance (Embrechts et al. 1997; Fontnouvelle et al. 2007; Abad and Benito 2013, among others). For instance, the extreme value of vehicle load plays an important role in bridge design and risk assessment (Wu and Qiu 2018). In seismology and climatology, extreme value analysis is used to study earthquakes (Beirlant et al. 2018) and extreme precipitation (Bader et al. 2018). In hydrology, extreme value analysis is an important tool for studying coastal flood risk (Haigh et al. 2010; McMillan et al. 2011). In the field of finance, extreme value modelling is important to quantify large financial losses from different sources of risk: operational, credit and market risk (see Cruz 2002; Moscadelli 2004; Fontnouvelle et al. 2007; Ergün and Jun 2010; Žiković and Aktan 2009; Abad and Benito 2013).

Traditionally, the study of the extreme values of fat tail distributions has been carried out through the extreme value theory (EVT). EVT comprises mainly two approaches—the block maximum method (BMM) and the POT approach—. In the former, the data set is divided into blocks and a generalized extreme value distribution is fitted to the sample of maximums or minimums extracted from these blocks. In the context of the POT method, a threshold is determined above which the excesses are fitted with the GPD (Queensley et al. 2019).

Although the BMM and POT approaches should lead asymptotically to the same results, in practice the POT provides more suitable extreme quantile estimations due to the more efficient use of the data for the extreme values, see Cunnane (1973) and Madsen et al. (1997a). These authors show that POT approach performs better than BMM, independently of the estimation method used. Similar results have been reported by Wang (1991), Madsen et al. (1997b).

From a theoretical point of view, threshold selection is a critical issue in the framework of the POT approach. The choice of threshold must be a balance between bias and variation. A threshold being too low is likely to violate the asymptotic basis of the model which leads to bias. However, a threshold being too high will generate few excesses leading to an increase in the variance of the estimators (Davison and Smith 1990; Coles 2001; MacDonald et al. 2011; Papalexiou et al. 2013; Wyncoll and Gouldby 2013).

That is why, within the framework of POT method, different methods have been developed for the selection of the suitable threshold. Those methods can be divided into two groups: (i) graphical approaches, based on a visual inspection of plots, such as the mean excess plot (Davison and Smith 1990), stability parameters plot (Coles 2001) and Hill plot (Drees et al. 2000) among others and (ii) numerical approaches (Ferreira et al. 2003; Thompson et al. 2009; Northrop and Coleman 2014; Li et al. 2014; Wadsworth and Tawn 2012; Naveau et al. 2016) which are more objective methods. Recently, new methods have been developed with the aim of automating some of the existing proposals, especially those based on visual data inspection, see for instance Wu and Qiu (2018), Bader et al. (2018), Caballero-Megido et al. (2018) and Queensley et al. (2019) among others.

The aforementioned papers focus on studying new methods for the selection of the optimal threshold, assuming that the estimates of the quantiles of the generalized Pareto are highly sensitive to the threshold choice in which case such efforts would be fully justified (Langousis et al. 2016). In the field of finance, the existing



literature on this issue is quite scarce, especially in the area of market risk management. As far as we know in this field, there are no studies on this subject. To cover this gap, we carry out an empirical analysis with a double aim. First, to analyse the sensibility of the GPD quantiles to the threshold choice and second, to study the sensibility of the market risk measure to this choice. For measuring market risk, we use the value at risk (VaR)¹ measure and the expected shortfall (ES)² measure. To last, we calculate the market risk capital requirements and evaluate their sensitive to the threshold choice.

This study is in accordance with Langousis et al. (2016) who remarked that “the variety of existing methods for threshold chosen, the fundamental differences in their theoretical underpinnings, and their relative performance when dealing with different types of data, make threshold detection an open question that can be addressed solely on the basis of a specific application”. With regard to this, we think, that in the area of market risk management, where daily data and large samples are used could give rise to different results from those obtained in other areas of science where the periodicity of the data is annual and consequently the size of the samples is reduced.

Thus, in this paper, we analyse in detail the case of the S&P500 and later extend that study to a set of 14 assets from alternative markets: seven stock indexes (CAC40, DAX30, FTSE100, HangSeng, IBEX35, Merval and Nikkei), four commodities (Copper, Gold, Crude Oil Brent and Silver) and three rates exchange (£/€, \$/€ and ¥/€). The results obtained indicate that there is a large set of thresholds that provide similar GPD quantiles estimators and as a consequence similar market risk measures. Just only, for large thresholds, those corresponding to the 98th and 99th percentile of the GPD some differences are found. It means that the choice of threshold in the framework of the POT method may not be relevant in quantifying market risk when we use the VaR and ES measures for this task. With regard to the market risk capital requirement, we find that these charges do not differ much among the thresholds. Nevertheless, if the objective of the financial institutions is to minimize these charges, they might be interested in the selection of a specific threshold.

The remainder of the paper is organized as follows. In “[Methodology](#)” section, we show the methodology used for the study. In “[Case study](#)” section, we submit the data and the results obtained for the particular case of the S&P500 index. “[Robustness analysis](#)” section displays a robustness analysis. The main conclusions are presented in “[Conclusions](#)” section.

¹ The VaR of a portfolio is defined as the worst expected loss over a given horizon under normal market conditions at a given level of confidence. Formally speaking, the $\text{VaR}(\alpha)$ of a portfolio at $(1 - \alpha)\%$ confidence level is the percentile $\alpha\%$ of the return portfolio distribution.

² The ES is defined as the average of all losses that are greater than or equal to VaR, i.e. the average loss in the worst $\alpha\%$ cases. In other words, this measure provides the expected value of an investment in the worst $\alpha\%$ of the cases.



Methodology

Extreme value theory

The Extreme Value Theory (EVT) studies the asymptotic behaviour of extreme values of a random variable. This theory has wide applications in many fields, such as, civil engineering (Wu and Qiu 2018), climatology (Davison and Smith 1990; Kharin et al. 2013), seismology (Beirlant et al. 2018), hydrology (Katz et al. 2002; Carreau et al. 2017; Bader et al. 2018), insurance (Reiss and Thomas 2007) and finance (Embrechts et al. 1997), among others.

Within the EVT context, there are two approaches that study extreme events. The first one, based on the Generalized Extreme Value (GEV) distribution, models the distribution of the minimum or maximum realizations and it is known as the Block Maxima (Minima) Method (BMM). The second one is the Peaks Over Threshold (POT) approach based on the Generalized Pareto distribution (GPD) (Pickands 1975) which models the exceedances over a particular threshold. In the next lines, we introduce these approaches.

Fisher–Tippett theorem

Suppose that X_1, X_2, \dots, X_n is a sequence of independently and identically distributed random variables with a distribution function $F(x) = \Pr(X_t \leq x)$ and denote $M_n = \max(X_1, X_2, \dots, X_n)$ as a sample of maxima of this series, the distribution function (CDF) of M_n is represented as

$$P(M_n \leq x) = P(X_1 \leq x, \dots, X_n \leq x) = \prod_{i=1}^n F(x) = F^n(x) \quad (1)$$

Although $F(x)$ is unknown, Fisher and Tippet theorem establishes an asymptotic approach for $F^n(x)$. This theorem establishes that given a sequence of $b_n > 0$, $a_n \in \mathbb{R}$, the maximum normalized value $Z_n = \frac{M_n - a_n}{b_n}$ converges to a non-degenerated distribution H , and this distribution is the generalized extreme value (GEV) distribution, $\lim_{n \rightarrow \infty} \Pr\left(\frac{M_n - a_n}{b_n} \leq x\right) \rightarrow H(x)$.

The algebraic expression for such generalized distribution is as follows:

$$\text{GEV}_{\xi, \mu, \sigma}(x) = e^{-\left[1 + \xi \frac{(x - \mu)}{\sigma}\right]^{-\frac{1}{\xi}}} \quad (2)$$

defined on $\left(1 + \frac{\xi(x - \mu)}{\sigma}\right) > 0$, where $\sigma > 0$ is the scale parameter, $-\infty < \mu < \infty$ is the mean, and $-\infty < \xi < \infty$ is known as the shape parameter of the GEV distribution and characterizes the tail behaviour of the distribution. The prior distribution is a generalization of the three types of distributions, depending on the value taken by ξ :

- Gumbel ($\xi = 0$) type I family. It has light extremes, not heavy extremes



$$\Lambda(x) = e^{e^{-\frac{x-\mu}{\sigma}}} \quad \forall x \in \mathfrak{R}$$

- Fréchet ($\xi > 0$) type II family. This distribution is particularly useful for patterning financial returns as it has very heavy tails.

$$\Phi_{\xi,\mu,\sigma}(x) = \begin{cases} 0 & x \leq \mu \\ e^{-\left(\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}} & x > \mu \end{cases}$$

- Weibull ($\xi < 0$) type III family. This distribution is used when the extremes are lighter (softer) than those from the normal distribution, and thus, it is not particularly useful for applications related to financial yields (returns).

$$\Psi_{\xi,\mu,\sigma}(x) = \begin{cases} e^{-\left(-\frac{x-\mu}{\sigma}\right)^{\frac{-1}{\xi}}} & x \leq \mu \\ 1 & x > \mu \end{cases}$$

Peaks over threshold approach (POT)

In general, we are not only interested in the maxima of observations, but also in the behaviour of large observations which exceed a high threshold. One method of extracting extremes from a sample of observations, $X_t, t = 1, 2, \dots, n$ with a distribution function $F(x) = \Pr(X_t \leq x)$ is to take the exceedances over a predetermined high threshold u . An exceedance of a threshold u occurs when $X_t > u$ for any t in $t = 1, 2, \dots, n$. Thus, an excess over u is defined as $y = X_t - u$. This approach is known as POT.

Although the BMM and POT approaches should lead asymptotically to the same results, in practice the POT provides more suitable extreme quantile estimations due to the more efficient use of the data for the extreme values, see Cunnane (1973) and Madsen et al. (1997a). These studies show that the POT approach performs better than BMM, independently of the estimation method used. Similar results have been reported by Wang (1991), Madsen et al. (1997b), and Tanaka and Takara (2002) among others.

Let x_0 be the finite or infinite right endpoint of the distribution F . That is to say, $x_0 = \sup\{x \in R : F(x) < 1\} \leq \infty$. The distribution function of the excesses (y) over the threshold u is given by $F_u(y) = P((X - u) \leq y | X > u)$ for $0 \leq y \leq x_0 - u$. Thus, $F_u(y)$ is the probability that the value of X exceeds the threshold u by no more than an amount y , given that the threshold is exceeded. This probability can be written as

$$F_u(y) = \frac{F(y + u) - F(u)}{1 - F(u)} \tag{3}$$

This distribution can be approximated by the generalized Pareto distribution (GPD) which is usually expressed as a two-parameter distribution³:

³ The study of Jobst (2007) provides favourable evidence in favour of this approach.



$$G_{k,\xi}(y) = \begin{cases} 1 - \left(1 + \frac{\xi}{\sigma}y\right)^{-\frac{1}{\xi}} & \text{if } \xi \neq 0 \\ 1 - \exp\left(-\frac{y}{\sigma}\right) & \text{if } \xi = 0 \end{cases} \tag{4}$$

where ξ and $\sigma > 0$ are the shape parameter and the scale parameter, respectively.⁴ Note that if the distribution of M_n^* converges to a GEV distribution for block maxima with parameter ξ , then the distribution of exceedances over threshold converges to the GPD with the same parameter ξ (Rodríguez 2017).

Using this approximation, the distribution function of X will be given by $F(x) = (1 - F(u))F_u(y) + F(u)$. Replacing $F_u(y)$ by GPD and $F(u)$ by its empirical estimator $(n - N_u)/n$, where n is the total number of observations and N_u the number of observations above the threshold u , we have

$$F(x) = 1 - \frac{N_u}{n} \left(1 + \frac{\xi}{\sigma}(x - u)\right)^{-\frac{1}{\xi}} \tag{5}$$

For a given probability $\alpha > F(u)$, the quantile α , which is denoted by q_α , is calculated by inverting the tail estimation formula to obtain

$$q_\alpha = u + \frac{\sigma}{\xi} \left(\left(\frac{n}{N_u}(1 - \alpha) \right)^{-\xi} - 1 \right) \tag{6}$$

The distributional choice is motivated by a theorem (Balkema and de Haan 1974; Pickands 1975) which states that, for a certain class of distributions, the GPD is the limiting distribution for the distribution of the excesses, as the threshold tends to the right endpoint:

$$\lim_{u \rightarrow x_0} \sup \left| F_n(y) - \text{GPD}_{\xi,\sigma}(y) \right| = 0$$

This theorem is fulfilled if and only if F is in the maximum domain of attraction of the generalized extreme value distribution H_ξ , ($F \in \text{MDA}(H_\xi)$). It means that if, for a given distribution F , an appropriately normalized maximum sample converges to a non-degenerated distribution H_ξ , then this is equivalent to say H_ξ is the MDA for F for some value of ξ .

⁴ The traditional extreme value theory (EVT) assumes that the data are stationary. When stationarity is assumed, parameters that determine the distribution function (generalized Pareto and generalized extreme value distribution) are independent of time. However, in practice, it is often the case that stationarity assumptions (such as independence and identical distribution) for time series extremes are violated. If the process is non-stationary, the parameters of distributions are time-dependent, and the properties of the distribution vary with time. To capture the non-stationarity of extreme data, new approaches have been developed in the framework of the extreme value theory. Some applications of these new approaches can be found in Cheng and AghaKouchak (2014), Cheng et al. (2014), Ruggiero et al. (2010), Chavez-Demoulin and Embrechts (2004) among others.



The class of distribution F for which the condition $F \in \text{MDA}(H_\xi)$ holds is large; essentially all commonly encountered continuous distributions show the kind of regular behaviour for sample maximum described by Eq. (1).

Threshold selection method

The approaches developed for selecting the suitable threshold can be divided into two groups: (i) subjective approaches based on graphical analysis, such as the mean excess plot, stability parameters plot and Hill plot among others and (ii) numerical approaches. In its turn, the latter can be divided into various categories: (a) non-parametric approach; (b) approaches based on goodness-of-fit test; (iii) mixture models; (iv) simple naïve methods; (v) computational approaches and (vi) other approaches. In the following lines, we describe briefly these methods (see Scarrot and McDonald 2012 and Langousis et al. 2016 for a more detailed review of these methods).

(i) Graphical approaches

Due to its simplicity, the graphic method most commonly used for selecting threshold is the mean excess plot (MEP) also called mean residual life plot (MRLP) introduced by Davison and Smith (1990). This instrument is a graphical tool based on the sample means of the excesses function (SMEF), which is defined as

$$\text{SMEF}(u) = \frac{\sum_{i=1}^{N_u} (r_i - u)_{\{r_i > u\}}}{N_u}$$

The sample means excess function is an estimate of the excess mean function (MEF), $e(u) = E[(X - u) | X > u]$. For the GPD, the excess mean function is given by a linear function in u ⁵:

$$e(u) = \frac{\sigma}{1 - \xi} + \frac{\xi}{1 - \xi} u \quad (7)$$

This finding means that for $0 < \xi < 1$ and $\sigma + u\xi > 0$, the mean excess plot should resemble a straight line with a positive slope. Empirical estimates of the sample mean excesses are typically plotted against a range of thresholds. Thus, the general rule for the choice of the optimal threshold will be to choose a value of u for which the resulting line has a positive slope. An application of this method can be found in Beirlant et al. (2004).

The main problem associated with the sample mean excess plot is subjectivity. As Queensley et al. (2019) remark, “judging from where the graph is approximately linear using only the eyeball inspection approach, is a rather subjective choice so that different thresholds may be selected by different viewers of the plot”.

⁵ If a distribution function is subexponential, the mean excess function tends to infinity, if it is an exponential distribution the mean excess function is a constant and for the normal distribution the mean excess function tends to zero.



The second type of plot is the parameter stability plot (Coles 2001) (shape and scale) created by fitting the GPD using a range of thresholds. This method involves plotting $\hat{\sigma}$ and $\hat{\xi}$ together with confidence intervals and selecting the value of u from which the estimates are no longer stable (see Coles 2001). This type of plot may present some inconsistencies, showing different flat sections for different ranges of threshold (Scarrot and McDonald 2012).

Other graphic approaches are these based on quantile plots and plots comparing the empirical cumulative distribution function and the cumulative GPD. According to quantile plots, the proper threshold is selected as the lowest threshold above which the plot shows a linear trend. When we compare the empirical with the theoretical distribution function, the proper threshold is selected as the lowest threshold above which the differences between the empirical and the theoretical distribution function seem minimum. Hill plot, explored by Drees et al. (2000) can also be included in this group. The Hill plot plots the Hill estimator of the tail index for a set of thresholds. According to this plot, the optimal threshold is the lowest threshold at which the Hill estimator is stabilised. This tool suffers from many of the same benefits and drawbacks that the MEP, and has been referred to as the Hill horror plot by Resnick (1997).

(ii) *Numerical approaches*

The approaches aforementioned are based on judgement (Caballero-Megido et al. 2018) so they can be rather subjective and require substantial expertise to interpret these diagnostics as a method of threshold selection (Davison and Smith 1990; Coles 2001; Solari and Losada 2012). To overcome these limitations, some numerical approaches have been developed which lead to a more objective decision. The numerical approaches are numerous and can be classified into different categories: (a) non-parametric approach; (b) approaches based on goodness-of-fit test; (c) simple naïve methods; (d) mixture models; (e) computational approaches and (e) other approaches. In the following lines, we resume each one of these categories.

- (a) *Nonparametric methods* that are intended to locate the changing point between extreme and nonextreme regions of the data (see e.g. Gerstengarbe and Werner 1989, 1991; Werner and Gerstengarbe 1997; Domonkos and Piotrowicz 1998; Lasch et al. 1999; Cebrián et al. 2003; Cebrián and Abaurrea 2006; Karpouzou et al. 2010, among others).
- (b) *Approaches based on the Goodness of fit test* where the threshold is selected as the lowest level above which the GPD provides an adequate fit to the exceedances. To analyse the goodness of fit of the GPD, Kolmogorov–Smirnov test and Anderson–Darling test can be used. Applications of this method can be found in Davison and Smith (1990), Dupuis (1999), Choulakian and Stephens (2001), Northrop and Coleman (2014), Langousis et al. (2016) among others.

In this category, we also include the method based on the Root Mean Error (RMSE) proposed by Li et al. (2014). The RMSE measures the difference between analytical and observed CDFs of exceedances for different thresholds. The threshold with the lowest RMSE is considered the best one.

- (c) *Simple naïve methods* Given the general order statistic convergence properties, various rules of thumb have been derived from the literature. Simple fixed quan-



tile rules, like the upper 10% rule of DuMouchel (1983). Ferreira et al. (2003) use the square root of the number of data (n) to specify the number of exceedances (N_u). Ho and Wan (2002) and Omran and McKenzie (2010) use the rule $N_u = \frac{n^{2/3}}{\log(\log(n))}$ proposed by Loretan and Philips (1994) to determine the optimal number of exceedances. Neftci (2000), followed by Bekiros and Georgoutsos (2005), proposes the estimation of the threshold as $1.176\sigma_0$, where σ_0 is the standard deviation of the sample. In other studies, these methods are classified as ad hoc methods or rules of thumb.

- (d) *Methods in the other category* are based on mixtures of a GPD for the tail and another distribution for the “bulk” joined at the threshold (e.g. MacDonald et al. 2011; Wadsworth and Tawn 2012; Naveau et al. 2016). Treating the threshold as a parameter to estimate, these methods can account for the uncertainty from threshold selection in inferences. The major drawback of such models is their ad hoc heuristic definitions, the asymptotic properties of which are still little understood. They have also not had time to be well used in practice and currently there is no readily available software implementation to allow practitioners to gain wider experience (Scarrot and McDonald 2012).
- (e) *Computational approaches* Other researchers have suggested using techniques that provide an optimal trade-off between bias and variance. This method involves using bootstrap simulations to numerically calculate the optimal threshold considering the trade-off between bias and variance. Applications of this method can be found in Danielsson et al. (2001), Drees et al. (2000), Ferreira et al. (2003), Hall (1990) and Beirlant et al. (2004). In general, the restrictive assumptions underlying these approaches hinder their wide applicability.
- (f) *Other approaches* Other approaches different from the aforementioned are proposed by Dupuis (1999), Thompson et al. (2009) and De Zea Bermudez et al. (2001). See Scarrot and McDonald (2012) for a detailed review of these methods.

Recently, new methods have been developed to automate some of the existing proposals, especially those based on visual data inspection, see for instance Wu and Qiu (2018), Bader et al. (2018), Caballero-Megido et al. (2018) and Queensley et al. (2019), Schneider et al. (2021) among others. Wu and Qiu (2018) propose a method to select the suitable threshold based on multiple criteria decision analysis (MCDA). In MCDA, Chi Square test, Kolmogorov–Smirnov (K-S) test and Root Mean Square Error (RMSE) are combined as the test criteria and the weight of these criteria is calculated using the entropy method. Thus, the MCDA can integrate results obtained from the goodness-of-fit test under different criteria into a comprehensive one, which makes the selection more scientific and objective (Wang et al. 2009). Bader et al. (2018) develop an efficient technique to evaluate and apply the Anderson–Darling test to the sample of exceedances above a fixed threshold. In order to automate threshold selection, this test is used in conjunction with a recently developed stopping rule that controls the false discovery rate in ordered hypothesis testing. Caballero-Megido et al. (2018) propose a new automated method that mimics the enduringly popular visual inspection method. The purpose of the automated graphic threshold selection (AGTS) method, in the absence of a priori threshold value, is to



guide in the choice of the threshold which requires judgement and expertise, making the process simple and approachable, whilst being reproducible and less subjective. Queensley et al. (2019) propose an alternative way of selecting the threshold where, instead of choosing individual thresholds in isolation and testing their fit, they make use of the bootstrap aggregate of these individual thresholds which are formulated in terms of quantiles. The method incorporates the visual technique and is aimed at reducing the subjectivity associated with solely using the eye inspection approach (EIA). Schneider et al. (2021) suggest a couple of automated methods for threshold selection. The first one consists in estimating and minimizing the integrated square error (ISE) between the exponential density and its parametric estimator employing the Hill estimator. This is based on the null hypothesis that the log-spacings between a sample of thresholds are indeed exponentially distributed. The error function that obtains is called the inverse Hill statistic (IHS). This method exhibits high fluctuations for small thresholds, which might make the automated selection of the minimum highly variable. To control this problem, the authors propose a smooth IHS. The second method consists in look for a sample fraction of optimal thresholds that minimise the asymptotic mean squared error (AMSE) of the Hill estimator.

Risk measure

According to Jorion (2001), “VaR measure is defined as the worst expected loss over a given horizon under normal market conditions at a given level of confidence”. Thus, VaR is a conditional quantile of the asset return loss distribution.

Let X_1, X_2, \dots, X_n be identically distributed independent random variables representing the financial returns. Using $F(x)$ to denote the cumulative distribution function, $F(x) = \Pr(X_t \leq x | \Omega_{t-1})$ conditioned to the information available at $t - 1$ (Ω_{t-1}). Assume that $\{X_t\}$ follows the stochastic process given by

$$X_t = \mu_t + \tilde{\sigma}_t z_t \quad z_t \sim iid(0, 1) \quad (8)$$

where $\tilde{\sigma}_t^2 = E(z_t^2 | \Omega_{t-1})$ and z_t has the conditional distribution function $G(z)$, $G(z) = P(z_t < z | \Omega_{t-1})$. The VaR with a given probability $\alpha \in (0, 1)$, denoted by $\text{VaR}(\alpha)$, is defined as the α quantile of the probability distribution of financial returns $F(\text{VaR}_t(\alpha)) = \Pr(X_t < \text{VaR}_t(\alpha)) = \alpha$. In this paper, we use POT approach to estimate the tail of the distribution of the standardized residuals and thus later estimate the risks measure. As the GPD is only defined for positive values, we multiply our data by (-1) and thus move the left tail to the right side. Therefore, the VaR of a portfolio at $\alpha\%$ probability will be calculated as

$$\text{VaR}_t(\alpha) = \mu_t + \tilde{\sigma}_t q_{1-\alpha} \quad (9)$$

where μ_t and $\tilde{\sigma}_t$ represent the conditional mean and the conditional standard deviation of the returns⁶ and $q_{1-\alpha}$ is the quantile $(1 - \alpha)$ of the GPD (Eq. 6).

⁶ For estimating the volatility of the return, we use an APARCH model, which is given by the next expression: $\sigma_t^\delta = \alpha_0 + \alpha_1 (|\epsilon_{t-1}| - \gamma \epsilon_{t-1})^\delta + \beta \sigma_{t-1}^\delta$, $\alpha_0, \beta, \delta > 0$, $\alpha_1 \geq 0$, $-1 < \gamma < 1$. In this model, the γ parameter captures the leverage effect (Black 1976), which means that volatility tends to be higher after negative returns.



The ES with a given probability $\alpha \in (0, 1)$, denoted by $ES(\alpha)$, is defined as the average of all losses that are greater than or equal to VaR , i.e. the average loss in the worst α % cases:

$$ES_t(\alpha) = E[X|X \geq VaR(\alpha)] = \mu_t + \tilde{\sigma}_t E[z|z \geq q_{1-\alpha}] \quad (10)$$

It can be demonstrated⁷ that the mean of the excess distribution $F_{q_{1-\alpha}}(y)$ over the threshold $q_{1-\alpha}$ is given by

$$E(z|z \geq q_{1-\alpha}) = \frac{q_{1-\alpha}}{1-\xi} + \frac{\sigma - \xi u}{1-\xi} \quad (11)$$

Replacing (11) in (10) we obtain the ES measure under the conditional EVT.

$$ES_t(\alpha) = E[X|X \geq VaR(\alpha)] = \mu_t + \tilde{\sigma}_t \left[\frac{q_{1-\alpha}}{1-\xi} + \frac{\sigma - \xi u}{1-\xi} \right] \quad (12)$$

Backtesting

Backtesting VaR

To evaluate the accuracy of the VaR estimates, several tests have been used. All of these tests are based on the indicator variable. We have an exception when $r_{t+1} < VaR_\alpha$; then, the exception indicator variable (I_{t+1}) is equal to one (zero in other cases).

To check the accuracy of the VaR estimates, we have used four standard tests: unconditional (LR_{uc}), independent and conditional coverage (LR_{ind} and LR_{cc}) and dynamic quantile (DQ) tests.

Kupiec (1995) shows that if we assume that the probability of obtaining an exception is constant, the number of exceptions $x = \sum I_{t+1}$ follows a binomial distribution $B(N, \alpha)$, where N represents the number of observations. An accurate measure VaR_α should produce an unconditional coverage $\left(\hat{\alpha} = \frac{\sum I_{t+1}}{N}\right)$ equal to α percent. The unconditional coverage test has a null hypothesis $\hat{\alpha} = \alpha$, with a likelihood ratio statistic:

$$LR_{uc} = 2 \left[\log(\hat{\alpha}^x (1 - \hat{\alpha})^{N-x}) - \log(\alpha^x (1 - \alpha)^{N-x}) \right] \quad (13)$$

which follows an asymptotic $\chi^2(1)$ distribution. The conditional coverage test, developed by Christoffersen (1998), jointly examines whether the percentage of exceptions is statistically equal to the one expected ($\hat{\alpha} = \alpha$) and the serial independence

⁷ A more detailed theoretical development can be found in McNeil et al. (2005), Chapter 7.



of the exception indicator. The likelihood ratio statistic of this test is given by $LR_{cc} = LR_{uc} + LR_{ind}$, which is asymptotically distributed as $\chi^2(2)$, and the LR_{ind} statistic is the likelihood ratio statistic for the hypothesis of the serial independence against first-order Markov dependence.⁸ Finally, the dynamic quantile test proposed by Engle and Manganelli (2004) examines if the exception indicator is uncorrelated with any variable that belongs to the information set Ω_{t-1} , available when the VaR is calculated. This test is a Wald test of the hypothesis that all slopes are zero in the regression:

$$I_t = \beta_0 + \sum_{i=1}^p \beta_i I_{t-i} + \sum_{j=1}^q \mu_j X_{t-j} \tag{14}$$

where X_{t-j} are the explanatory variables contained in Ω_{t-1} . This statistic is introduced as five explanatory variable lags of VaR. Under the null hypothesis, the exception indicator cannot be explained by the level of VaR, i.e. $VaR(\alpha)$ is usually an explanatory variable to test if the probability of an exception depends on the level of the VaR.

Backtesting ES

In this paper, we use McNeil and Frey (2000) test for the conditional expected shortfall. This test is likely the most successful in the literature. These authors develop a test to verify that a model provides much better estimates of the conditional expected shortfall than any other. The authors are interested in the size of the discrepancy between the return r_{t+1} and the conditional expected shortfall forecast $ES_t(\alpha)$ in the event of quantile violation. The authors define the residuals as follows:

$$Y_{t+1} = \frac{r_{t+1} - ES_{t+1}(\alpha)}{\sigma_{t+1}} \tag{15}$$

Replacing Eqs. (8) and (12) in Eq. (15), we have the next expression:

$$y_{t+1} = z_{t+1} - E(z|z < q_\alpha) \tag{16}$$

It is clear that, under model (5), these residuals are *i.i.d.* and that, conditional on $\{r_{t+1} < VaR_{t+1}(\alpha)\}$ or equivalent $\{z_{t+1} < q_\alpha\}$, they have an expected value of zero. Suppose we again backtest on days in the set T . We can form empirical versions of these residuals on those specific days on which violations have occurred, i.e. days in which $\{r_{t+1} < VaR_{t+1}(\alpha)\}$. The authors call these residuals exceedances and denote them by $\{\hat{y}_{t+1} : t \in T, r_{t+1} < VaR_{t+1}(\alpha)\}$ where $\hat{y}_{t+1} = \frac{r_{t+1} - \widehat{ES}_{t+1}(\alpha)}{\hat{\sigma}_{t+1}}$ and $\widehat{ES}_{t+1}(\alpha)$ is an estimation of the conditional expected shortfall.

⁸ The LR_{ind} statistic is $LR_{ind} = 2[\log L_A - \log L_0]$ and has an asymptotic $\chi^2(1)$ distribution. The likelihood function under the alternative hypothesis is $L_A = (1 - \pi_{01})^{N_{00}} \pi_{01}^{N_{01}} (1 - \pi_{11})^{N_{10}} \pi_{11}^{N_{11}}$, where N_{ij} denotes the number of observations in state j after having been in state i in the previous period, $\pi_{01} = N_{01}/(N_{00} + N_{01})$ and $\pi_{11} = N_{11}/(N_{10} + N_{11})$. The likelihood function under the null hypothesis ($\pi_{01} = \pi_{11} = \pi = (N_{11} + N_{01})/N$) is $L_0 = (1 - \pi)^{N_{00} + N_{01}} \pi^{N_{01} + N_{11}}$.



Under the null hypothesis, in which we correctly estimate the dynamic of the process μ_{t+1} and σ_{t+1} and the first moment of the truncated innovation distribution $E(z|z < q_\alpha)$, these residuals should behave such as an *i.i.d* sample with a mean of zero. Thus, for testing whether the estimates of the expected shortfall are correct, we must test if the sample mean of the residual is equal to zero against the alternative that the mean of y is negative. Indeed, given a sample $\{y_{t+1}\}$ of size N (where N is the number of violations in the T period), the sample mean \bar{y} converges in distribution to standard normality, as N tends to ∞ by the central limit theorem. In other words, given mean μ_y and variance σ_y of population

$$\sqrt{N} \left(\frac{\bar{y} - \mu_y}{\sigma_y} \right) \rightarrow N(0, 1) \quad (17)$$

By applying the central limit theorem, the statistics for testing the null hypothesis are given by

$$t = \frac{\bar{y}}{\frac{S_y}{\sqrt{N}}} \sim t_{N-1} \quad (18)$$

where \bar{y} and S_y are the sample mean and the sample standard deviation, respectively, of the exceedance residuals.

Forecasting daily market risk capital charges

Basel II Accord required financial institutions to meet daily capital requirements based on VaR estimates (BCBS 1996, 2006). The Basel II Accord specified that daily capital charges (DCC) must be set at the higher of the previous day's VaR or the average VaR over the last 60 business days, multiplied by a value between 3 and 4 depending on the number of violations (see Table 1) that occurred in the 250 days prior to the estimation of capital charges $DCC_t = \sup\{-k \times \overline{\text{VaR}}_{60}, -\text{VaR}_{t-1}\}$.

Recently, the Basel Committee for Banking Supervision (BCBS) has promoted a change in international financial regulation. Under the new regulation based on the Basel solvency framework (BCBS 2012, 2016, 2017, 2019), known as Basel III, financial institutions must calculate the market risk capital requirements based on the Expected Shortfall (ES) measure, replacing the Value at Risk (VaR) measure.

Following Chang et al. (2019), we evaluate the market risk capital requirement based on ES measure, which is the market risk benchmark according to Basel III. Thus, the forecasting daily market risk capital requirement (DCR) at time t can be calculated as follows:

$$\text{DCR}_t = \sup\left\{-k \times \overline{\text{ES}}_{60}, -\text{ES}_{t-1}\right\} \quad (19)$$



Table 1 Basel accord penalty zones

Zone	Number of exceptions	k
Green	0 to 4	3
Yellow	5	3.4
	6	3.5
	7	3.65
	8	3.75
	9	3.85
Red	10 or more	4

The number of exceptions is given for 250 trading days

Case study

Dataset overview

The data consist of the S&P500 stock index extracted from the Thomson-Reuters-Eikon database. The index is transformed into returns by taking the logarithmic differences of the closing daily price (in percentage). We use daily data for the period January 3rd, 2000, through December 30th, 2021. The sample size is 5534. Figure 1 shows the evolution of the daily index and returns of the S&P500. The index shows a sawtooth profile alternating periods with an upward slope with a period of sudden decreases.

In addition, we can observe that the range fluctuation of daily returns is not constant, which means that the variance of the returns changes over time. The volatility of S&P500 was particularly high from 2008 to 2009, coinciding with the period known as the Global Financial Crisis, and in the first quarter of 2020, coinciding with the beginning of the COVID-19 pandemic. The basis descriptive statistics are provided in Table 2. The unconditional mean daily return is very close to zero

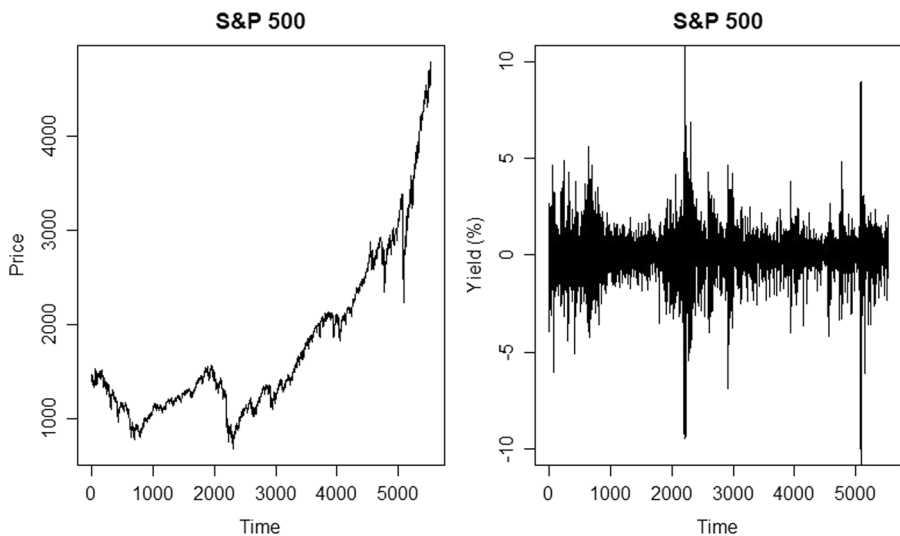


Fig. 1 S&P500



Table 2 Descriptive statistics

	Mean	Median	Maximum	Minimum	SD	Skewness	Kurtosis	Jarque Bera
S&P 500	0.0214	0.0639	10.957	- 12.765	1.2391	- 0.401* (0.033)	11.067* (0.066)	28,334 (0.000)

This table presents the descriptive statistics of the daily returns of S&P500. The sample period is from January 3rd, 2000 to December 30th, 2021. The index return is calculated as $R_t = 100(\ln(P_t) - \ln(P_{t-1}))$ where P_t is the index level for period t . Standard errors of the skewness and excess kurtosis are calculated as $\sqrt{6/n}$ and $\sqrt{24/n}$, respectively. The JB statistic is distributed as the Chi square with two degrees of freedom

*Significance at the 5% level

(0.021%) which is typical of daily returns. The skewness statistic is negative, implying that the distribution of daily returns is skewed to the left. The kurtosis coefficient shows that the distribution has much thicker tails than the normal distribution. Similarly, the Jarque-Bera statistic is statistically significant, rejecting the assumption of normality. All this evidence shows that the empirical distribution of daily returns cannot be fit by a normal distribution, as it exhibits a significant excess of kurtosis and asymmetry (fat tails and peakness).

Before continuing, we briefly summarize the steps performed in this study. For each of the thresholds selected, *first of all*, we evaluate the fit of the GPD, *second*, we analysed the stability of the GPD parameters. In the *third* place, the sensitivity of the high quantiles of the GPD to the threshold choice is evaluated (“[Fitting the GPD](#)” section). *Later*, we evaluate the sensitivity of the risk market measure to the threshold choice (“[Sensitivity of the risk measures to changes in the threshold](#)” section). *Fifth*, we assess the accuracy of the estimated risk measures (“[Analysing the quality of the risk estimates](#)” section). *Finally*, for each selected threshold, we calculate the capital charges based on the ES measure (“[Analysing the sensitivity of forecasting daily capital charges to the selected threshold](#)” section). The objective is to evaluate how sensitive the capital requirements are to the choice of the threshold.

Fitting the GPD

In this section, we fit GPD to the data for a set of 20 thresholds. This section aims to evaluate the sensitivity of the parameters and quantiles of the generalized Pareto distribution (GPD) to changes in the threshold. The thresholds were chosen from a quantile range between the 80th and the 99th percentile at 1% increments. As the size of the sample is 5534 daily returns, the percentile 80th gives 1107 exceedances, whilst the 99th percentile gives 56 exceedances. As the threshold increases by one unit, the number of exceedances decreases by 55 units. According to the theory, the distribution of the exceedances defined as $F_n(X - u)$ for $X \geq u$ may be approximated by the GPD denoted by $G_{\xi, \sigma}(y)$. Thus, for each threshold, we fit a GPD and check that the sample of the excesses above the threshold follows a $G_{\xi, \sigma}(y)$. Examples of this fit can be seen in Fig. 2 for the 5534 returns with thresholds set at 0.74% and 2.76%. These thresholds give 1107 and 56 exceedances, respectively. Parameters are estimated by maximum likelihood and the resulting GPD curves are superimposed



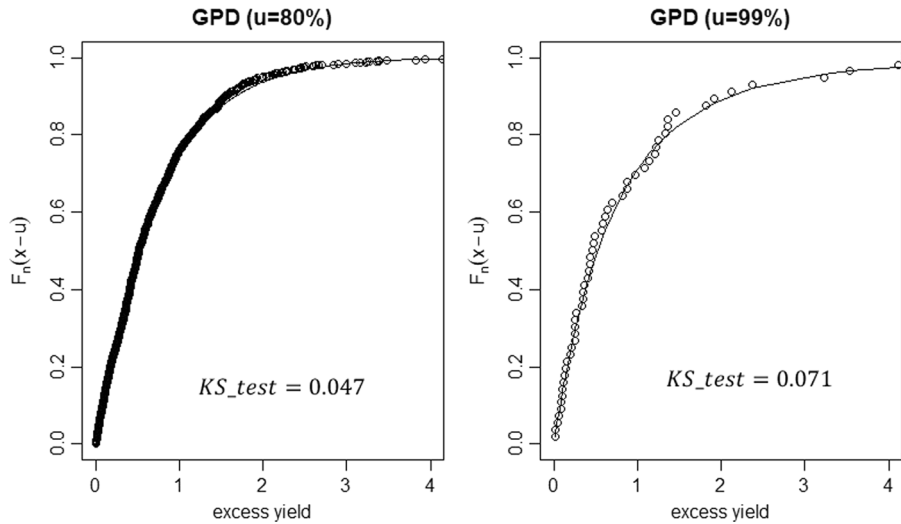


Fig. 2 GPD. In the left plot, GPD is fitted to 1107 exceedances over the threshold of 0.74%. In the right plot, GPD is fitted to 56 exceedances over the threshold of 2.76%

on the empirical estimate of the distribution function of the exceedances. As we can see, GPD seems to fit pretty well the exceedances samples. In concordance with this, the Kolmogorov–Smirnov test used to test if the sample of exceedances follows a GPD cannot be rejected in any case (see Table 3).

Let u_1, u_2, \dots, u_n be the set of thresholds selected ($n = 20$). For $j = 1, \dots, n$, let $\hat{\xi}_{u_j}$ and $\hat{\sigma}_{u_j}$ be the estimators of the shape and scale parameters based on the exceedances over the threshold u_j . Figure 3 displays the estimation of ξ and σ , respectively, as a function of the threshold u .

We observe that as the threshold increases, the value of ξ increases. In the case of the scale parameter, the opposite occurs; as the threshold increases, the value of σ is reduced. As we expected, in both cases, the accuracy of the estimations decreases as the threshold increases.

The estimation of the shape parameter, which determines the weight of the tail in the distribution, is very sensitive to changes in the threshold. For instance, the value of ξ increases by 2233% when the threshold moves from the 80th percentile to the 90th percentile. From the 90th percentile to the 99th percentile, the increase is equal to 227%. The value of the scale parameter is also sensitive to changes in the threshold; however, in this case, the changes are not that striking. Thus, in accordance with the literature, we find that the parameter estimations are very sensitive to the threshold selected for estimating GPD. But, what about the GPD quantiles? Do they depend on the threshold choice? To answer this question, we analyse the sensitivity of the high quantiles generalized Pareto distribution to changes in the threshold. For this analysis, we just only focus on the high quantiles (95%, 96%, 97%, 98% and 99%) as they are the only relevant quantiles in quantifying market risk. The quantile of the GPD is calculated using the Eq. (6).



Table 3 Maximum likelihood estimations (GPD)

Percentiles	Threshold return	Exceedances	ξ	σ	KS test	Critical value
80th	0.74	1107	0.003 (0.024)	0.722 (0.028)	0.047	0.041*
81th	0.78	1052	0.006 (0.025)	0.718 (0.028)	0.048	0.042*
82th	0.82	996	0.009 (0.025)	0.714 (0.029)	0.050	0.043*
83th	0.87	941	0.013 (0.026)	0.706 (0.029)	0.051	0.044*
84th	0.92	886	0.020 (0.028)	0.695 (0.030)	0.048	0.046*
85th	0.98	830	0.042 (0.031)	0.660 (0.030)	0.033	0.047
86th	1.04	775	0.063 (0.034)	0.633 (0.031)	0.019	0.049
87th	1.10	720	0.079 (0.037)	0.614 (0.032)	0.023	0.051
88th	1.15	664	0.075 (0.038)	0.623 (0.033)	0.024	0.053
89th	1.20	609	0.080 (0.040)	0.621 (0.035)	0.026	0.055
90th	1.26	554	0.070 (0.040)	0.639 (0.037)	0.030	0.058
91th	1.32	498	0.073 (0.042)	0.641 (0.039)	0.034	0.061
92th	1.41	443	0.087 (0.046)	0.626 (0.041)	0.031	0.065
93th	1.51	388	0.105 (0.051)	0.611 (0.043)	0.028	0.070
94th	1.62	332	0.135 (0.058)	0.585 (0.046)	0.039	0.075
95th	1.72	277	0.124 (0.061)	0.613 (0.052)	0.046	0.082
96th	1.88	222	0.173 (0.075)	0.572 (0.057)	0.058	0.091
97th	2.07	166	0.206 (0.091)	0.567 (0.067)	0.072	0.106
98th	2.28	111	0.207 (0.115)	0.628 (0.093)	0.045	0.129
99th	2.76	56	0.229 (0.171)	0.696 (0.149)	0.071	0.182

ξ Shape parameter, σ scale parameter. The standard deviation is given in parenthesis. KS test is the Kolmogorov–Smirnov test. The critical value at 95% of confidence level is calculated as $1.36/\sqrt{(n)}$. In the cases denoted with (*), we cannot reject the null hypothesis at 1%. The critical value at 99% of confidence level is calculated as $1.63/\sqrt{(n)}$

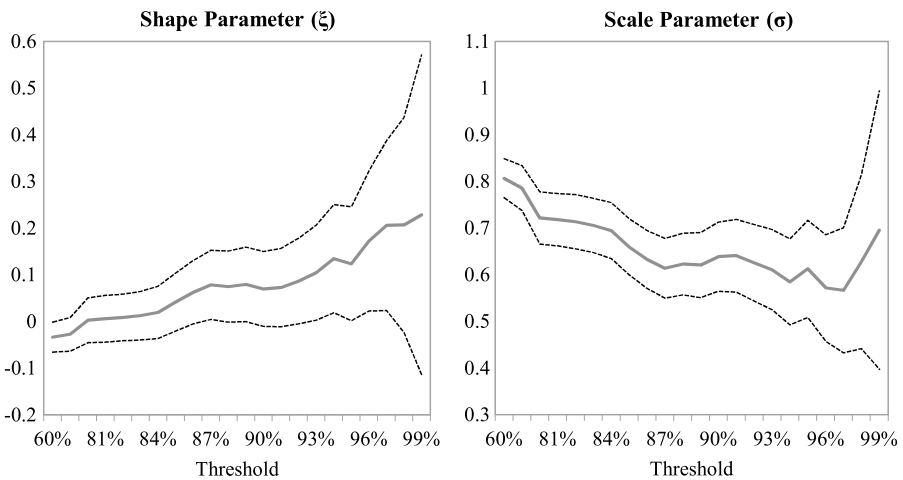


Fig. 3 Maximum Likelihood Estimations GPD



Table 4 Differences in quantiles

Panel (a)						Panel (b)					
u	Differences in quantiles					u	Differences in quantiles				
	95%	96%	97%	98%	99%		95%	96%	97%	98%	99%
80 th	3	4	6	6	6	80 th	-5	0	5	11	15
81 st	3	4	5	6	6	81 st	-5	0	5	10	15
82 nd	3	4	5	6	6	82 nd	-5	0	5	10	15
83 rd	3	4	5	5	5	83 rd	-5	-1	4	9	14
84 th	2	3	4	5	4	84 th	-6	-1	3	9	13
85 th	1	2	2	2	2	85 th	-7	-3	2	7	11
86 th	0	1	1	1	0	86 th	-7	-4	0	5	9
87 th	0	0	-1	-1	-1	87 th	-8	-5	-1	3	8
88 th	0	0	0	0	0	88 th	-8	-5	-1	4	9
89 th	0	0	-1	-1	-1	89 th	-8	-5	-1	3	8
90th	0	0	0	0	0	90 th	-8	-4	0	4	9
91 st	0	0	0	0	0	91 st	-8	-5	-1	4	9
92 nd	0	0	0	0	0	92 nd	-8	-5	-1	3	8
93 rd	0	0	-1	-1	-1	93 rd	-7	-5	-2	2	6
94 th	0	0	-1	-2	-3	94 th	-7	-5	-2	1	4
95 th	1	0	-2	-4	-5	95 th	-7	-5	-2	1	5
96 th	1	0	-2	-3	-4	96 th	-3	-2	-1	0	2
97 th	5	2	-1	-4	-7	97th	0	0	0	0	0
98 th	8	4	0	-4	-9	98 th	-3	-3	-2	-2	-1
99 th	5	2	-2	-6	-10	99 th	4	3	3	2	0

In the left plot the GPD quantiles at 95%, 96%, 97%, 98%, and 99% probability are displayed as a function of the thresholds. The tables capture the differences in quantiles related to the 90th (Panel (a)) and 97th percentile (Panel (b)). We shaded the differences that oscillate between 3 and 4 basis points in light gray. Differences greater than 4 basis points are shaded in dark gray

Figure 4 displays these quantiles as a function of the threshold u . What pays our attention is that the line representing the quantiles as a function of the threshold is completely flat, which means that the high quantile of the GPD does not depend on the threshold choice, at least in the range of threshold considered in this paper. This result is pretty striking. Table 4 displays the differences in α quantiles obtained for all thresholds considered against the threshold benchmark. Panel (a) displays these differences regard to the threshold corresponding to percentile 90th, which is in the middle point of the considered range vector and has been used successfully in many empirical papers in VaR estimate (see Abad and Benito 2013; Benito et al. 2017). Panel (b) displays these differences with regard to the threshold corresponding to percentile 97th, which is the optimal threshold according to the excess mean plot.⁹ In Panel (a), we observe that for a large set of thresholds, from a return corresponding to the 81st percentile to a return corresponding to the 97th percentile, the differences in quantile estimation do not exceed the 7 basis points. Even more, from a return corresponding to the 85th percentile to a return corresponding to the 94th percentile, the differences in quantile estimation {95.0th to 98th} do not exceed the 2 basis points. As regards Panel (b), although the differences in quantile estimation

⁹ We do not present this plot for reason of space.



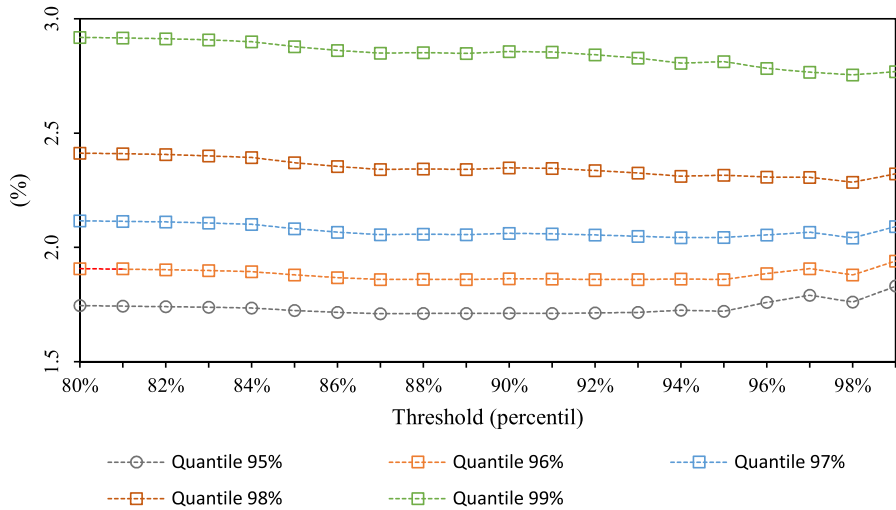


Fig. 4 GPD quantiles. GPD quantiles at 95%, 96% 97%, 98% and 99% probability are displayed as a function of the thresholds

{95.0th to 98th} are larger compared with Panel (a), they do not exceed 8 basis points from a return corresponding to the 80th percentile to a return corresponding to the 99th percentile. Just only in the case of the 99th quantile, the differences are somewhat higher.

As the estimation of the market risk depends on the quantile of the GPD, this preliminary analysis may suggest that the choice of the threshold in the framework of the POT method may not be very relevant in quantifying market risk.

Sensitivity of the risk measures to changes in the threshold

We can say that the analysis presented in the previous section is in accordance with the literature; we observe that the estimates of the parameters that describe the generalized Pareto distribution depend significantly on the threshold selected for the estimation. However, surprisingly the high quantiles of the GPD keep approximately constant. In this section, we want to go a step further by assessing to what extent the selection of the threshold affects the quantification of financial risk. With this objective, a set of 20 thresholds has been selected. The parametric estimates corresponding to these thresholds were presented in the previous section.

To quantify the risk, we use VaR and ES measures, which were presented in “Risk measure” section. The expression for these measures is given by

$$VaR_t(\alpha) = \mu_t + \tilde{\sigma}_t q_{1-\alpha} \quad ES_t(\alpha) = \mu_t + \tilde{\sigma}_t \left[\frac{q_{1-\alpha}}{1-\xi} + \frac{\sigma + \xi u}{1-\xi} \right] \quad (19)$$



where μ_t is the conditional mean return that is assumed constant ($\mu_t = \mu$), $\tilde{\sigma}_t$ represents the conditional standard deviation of the return; $q_{1-\alpha}$ is the percentile $1 - \alpha$ of the GPD and ξ and σ are the shape and scale parameters of the GPD. For the estimation of the conditional standard deviation of the returns, we use an APARCH model.

The sample period is divided into a learning sample from January 3rd, 2000, to December 30th, 2016, and a forecast sample from January 3rd, 2017, to the end of December 2021. For each day of the forecast period, we will generate estimations of VaR and ES measures. These forecasting measures are obtained one day ahead at the 95% and 99% confidence levels. Table 5 presents the descriptive statistics of the differences between the market risk estimates obtained from the threshold corresponding to the 90th percentile and the market risk estimates obtained from the remaining selected thresholds. For VaR estimates at the 95% confidence level, from a return corresponding to the 80th percentile to a return corresponding to the 96th percentile, the mean of the differences does not exceed the 3 basis points with a standard deviation between 1 and 3 basis points. For the thresholds corresponding to the 97th and 99th percentiles, the mean of the differences in the VaR estimate at 95% confidence level increases moving between 6 and 9 basis points.

The standard deviation of these differences also increases, moving between 4 and 28 basis points. For these thresholds, the minimum difference becomes 45 basis points (99th percentile), whilst the maximum difference becomes 229 basis points (99th percentile). For VaR estimates at the 99% confidence level, we find similar results. For a large set of thresholds (from the 82nd percentile to the 96th percentile), the mean and standard deviation of the differences are very small, not exceeding 5 basis points. Only in the case of the threshold corresponding to 99th percentiles,

Table 5 Differences between VaR and ES estimates

Threshold (u)	95% confidence level								99% confidence level							
	VaR				ES				VaR				ES			
	Mean	S.D.	Max	Min	Mean	S.D.	Max	Min	Mean	S.D.	Max	Min	Mean	S.D.	Max	Min
80%	-3	2	-1	-23	-4	3	-1	-24	-5	4	-2	-33	1	1	8	-1
81%	-3	2	-1	-22	-3	2	-1	-23	-5	3	-1	-32	1	1	8	0
82%	-3	2	-1	-19	-3	2	-1	-20	-4	3	-1	-27	1	1	8	0
83%	-2	2	-1	-18	-3	2	-1	-18	-4	3	-1	-25	1	1	9	0
84%	-2	2	-1	-16	-2	2	-1	-15	-3	2	-1	-21	1	1	9	0
85%	-1	1	0	-8	-1	1	0	-5	-2	1	0	-6	1	1	8	0
86%	0	0	0	-2	0	0	2	-1	0	1	6	-1	0	0	3	0
87%	0	0	1	0	0	0	5	0	1	1	10	0	0	0	0	-2
88%	0	0	0	0	0	0	4	0	0	1	8	-1	0	0	0	-1
89%	0	0	2	0	0	1	7	-1	0	2	15	-2	0	0	1	-3
90%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
91%	0	0	1	0	0	0	4	0	1	1	9	-1	0	0	0	0
92%	0	0	0	0	1	1	6	0	1	1	13	-1	0	0	0	0
93%	-1	0	0	-3	1	1	8	0	3	2	21	1	0	0	1	-1
94%	-1	1	0	-7	1	1	10	0	4	3	31	1	0	0	1	-1
95%	0	0	2	-1	1	1	9	0	3	2	20	0	0	0	1	-2
96%	-3	3	0	-28	1	1	5	0	5	4	40	1	0	1	6	0
97%	-6	4	-2	-38	0	1	3	-2	6	5	45	2	1	1	7	0
98%	-8	4	0	-27	-1	2	14	-5	7	5	50	2	1	1	5	-1
99%	9	28	229	-45	7	16	140	-19	8	9	84	2	0	1	2	-3

Note: This table shows some descriptive statistics of the differences between the VaR and ES estimations obtained under the threshold u_j ($j=1, \dots, 20$) and the VaR and ES estimates obtained under the optimal threshold. The optimal threshold is given by the 90th percentile. We shaded in light gray the differences that oscillate between 3 and 4 basis points. Differences greater than 4 basis points are shaded in dark gray.

Descriptive statistics. Optimal threshold 90%



the differences are higher. In summary, we find that for a large set of thresholds (the return corresponding to the 80th percentile to the 96th percentile), the quantification of risk that we obtain from VaR measures is similar. The same conclusion can be drawn from the ES measure. At the 95% confidence level and from the 82nd percentile to the 98th percentile, the mean and standard deviation of the differences do not exceed 3 basis points. At the 99% confidence level, the differences are even smaller and do not exceed 1 basis point (from the 80th percentile to the 99th percentile). Thus, we can conclude that in the selected range, the choice of threshold in the framework of the POT method may not be very relevant in quantifying market risk.

Analysing the quality of the risk estimates

In this section, we are interested in analysing the accuracy of the risk measures (VaR and ES) obtained from the conditional EVT. In addition, we will analyse if the quality of these measures depends on the threshold selected for applying EVT. Therefore, we will use the backtesting techniques presented in “Backtesting” section.

To evaluate the accuracy of the VaR estimates, we have used four standard tests: unconditional (LR_{uc}), independent (LR_{ind}), conditional coverage (LR_{cc}) and dynamic quantile (DQ) tests. The results of these tests are presented in Table 6. In this table, we also present the number and the percentage of exceptions. The first thing that pays our attention when viewing Table 6 is that for a large set of thresholds (from the 82nd percentile to the 93rd percentile), the number of exceptions is very close to the expected one.¹⁰ In the cases in which the number of exceptions differs from the theoretical one, the differences are very small. Thus, at the 95% confidence level, the percentage of exceptions ranges from 4.61% to 5.56%, corresponding to the 80th percentile and the 99th percentile. At the 99% confidence level, the percentage of exceptions ranges from 1.43% to 1.59%, also very similar to the expected one (1%). To test statistically whether the number of exceptions is equal to the theoretical one, we use the aforementioned test. We cannot reject the null hypothesis “that the VaR estimates are accurate” for any of the thresholds selected. To test whether the ES estimations are correct, we use the procedure proposed by McNeil and Frey (2000) test. The results of these tests are displayed in Table 6. The null hypothesis that states that the ES(95%) estimates are correct is rejected for all thresholds at both 5% and 1% probability. However, the hypothesis that the ES(99%) estimates are correct is rejected at 5% for all thresholds, but not at 1%.

The results presented in this section indicate that the choice of threshold in the framework of the POT method may not be relevant in quantifying market risk when we use the VaR and ES measures for this task.

¹⁰ For the forecasting period considered in this study, which has 1258 observations, the expected number of exceptions is 63 at a 95% confidence level and 13 at a 99% confidence level.



Table 6 Backtesting VaR and ES for S&P500 (2017–2021)

Threshold	95% confidence level							99% confidence level								
	Excep- tions		VaR				ES		Excep- tions		VaR				ES	
	No.	%	LR _{uc}	LR _{ind}	LR _{cc}	DQ	MF	No.	%	LR _{uc}	LR _{ind}	LR _{cc}	DQ	MF		
80th	58	4.61	0.67	0.38	0.62	0.01	0.01	18	1.43	0.34	0.45	0.48	0.18	0.03		
81st	58	4.61	0.67	0.38	0.62	0.01	0.01	18	1.43	0.34	0.45	0.48	0.18	0.03		
82nd	58	4.61	0.67	0.38	0.62	0.01	0.01	18	1.43	0.34	0.45	0.48	0.18	0.03		
83rd	58	4.61	0.67	0.38	0.62	0.01	0.01	18	1.43	0.34	0.45	0.48	0.18	0.03		
84th	58	4.61	0.67	0.38	0.62	0.01	0.00	18	1.43	0.34	0.45	0.48	0.18	0.03		
85th	58	4.61	0.67	0.38	0.62	0.01	0.00	18	1.43	0.34	0.45	0.48	0.18	0.03		
86th	58	4.61	0.67	0.38	0.62	0.01	0.00	18	1.43	0.34	0.45	0.48	0.18	0.03		
87th	59	4.69	0.74	0.24	0.47	0.01	0.00	18	1.43	0.34	0.45	0.48	0.18	0.04		
88th	59	4.69	0.74	0.24	0.47	0.01	0.00	18	1.43	0.34	0.45	0.48	0.18	0.03		
89th	59	4.69	0.74	0.24	0.47	0.01	0.00	18	1.43	0.34	0.45	0.48	0.18	0.04		
90th	59	4.69	0.74	0.24	0.47	0.01	0.00	18	1.43	0.34	0.45	0.48	0.18	0.03		
91st	59	4.69	0.74	0.24	0.47	0.01	0.00	18	1.43	0.34	0.45	0.48	0.18	0.03		
92nd	59	4.69	0.74	0.24	0.47	0.01	0.00	18	1.43	0.34	0.45	0.48	0.18	0.03		
93rd	58	4.61	0.67	0.38	0.62	0.01	0.00	18	1.43	0.34	0.45	0.48	0.18	0.03		
94th	58	4.61	0.67	0.38	0.62	0.01	0.00	19	1.51	0.26	0.48	0.42	0.23	0.05		
95th	59	4.69	0.74	0.24	0.47	0.01	0.00	18	1.43	0.34	0.45	0.48	0.18	0.04		
96th	58	4.61	0.67	0.38	0.62	0.01	0.00	19	1.51	0.26	0.48	0.42	0.23	0.05		
97th	56	4.45	0.55	0.54	0.69	0.01	0.00	19	1.51	0.26	0.48	0.42	0.23	0.05		
98th	55	4.37	0.49	0.51	0.64	0.00	0.00	19	1.51	0.26	0.48	0.42	0.23	0.05		
99th	70	5.56	0.55	0.32	0.51	0.00	0.07	20	1.59	0.20	0.52	0.36	0.28	0.08		

Analysing the sensitivity of forecasting daily capital charges to the selected threshold

In this section, we carry out an empirical application in which we evaluate the sensitivity of the daily market risk capital requirement (DCR) to the threshold choice. For this proposal, we follow Chang et al. (2019) and calculate the DCR according to Eq. (19). Figure 5 shows the mean of the DRC calculated on the base of ES measure at 99% confidence level for each of the thresholds selected.

The visual inspection of this figure suggests that there is a large set of thresholds (85th to 95th) that provide similar results, observing some differences in the lowest (80th to 84th) and highest (96th to 99th) thresholds. Table 7 which shows the mean, standard deviation and range of daily capital requirements confirms these previous results. For a large set of thresholds (80th to 95th proxy), the differences in DRC are under 5 basis points. This implies that for investment portfolios worth 1 million euros, the differences do not exceed 50 thousand euros. However, for thresholds outside this range, the differences are somewhat greater.



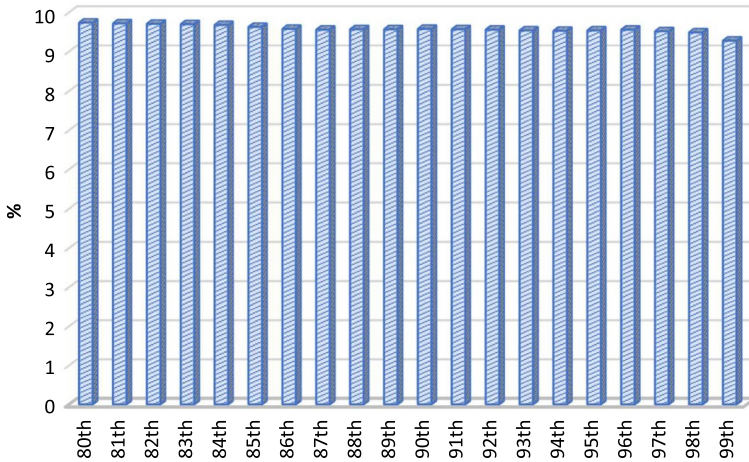


Fig. 5 Mean Market Risk Capital Requirement calculated on the ES measure

Table 7 Statistics of the daily capital requirement (ES 99%)

	Mean	S.D.	Range
80%	11,09	6,07	31,11
81%	11,09	6,07	31,12
82%	11,09	6,07	31,11
83%	11,09	6,07	31,11
84%	11,09	6,07	31,11
85%	11,09	6,07	31,13
86%	11,12	6,08	31,19
87%	11,14	6,09	31,23
88%	11,13	6,09	31,22
89%	11,14	6,10	31,26
90%	11,13	6,09	31,22
91%	11,13	6,09	31,23
92%	11,13	6,09	31,22
93%	11,13	6,09	31,23
94%	11,25	6,05	31,23
95%	11,14	6,10	31,24
96%	11,23	6,04	31,17
97%	11,22	6,02	31,16
98%	11,22	6,01	31,17
99%	11,45	6,50	33,96

Note: The table shows the mean of forecasting DCR for S&P500. The shadow cells report the mean DCR with a difference of less than 5 basis points

Robustness analysis

In the above section, we show that in the framework of the POT approach, there is a set of thresholds that provide a similar market risk estimate. It is due to the fact that the GPD quantiles are not sensitive to the threshold choice. Just only for the threshold 99th some differences are found. To corroborate the validity of this



result, in this section, we extend the study to a set of 14 assets. In accordance with the performed study for the S&P500, for each of these assets, we select a set of 20 thresholds, from 80th percentile to 99th percentile. For each threshold selected, first, we apply the conditional POT approach to analyse the sensitivity of the GPD quantiles to the threshold choice. Second, we obtain forecast VaR and ES measures 1 day ahead and analyse the differences among them for the set of thresholds selected. To last, we analyse the sensitive of the market risk capital charges to the threshold choice.

Before analysing the accuracy of the market risk measure, we evaluate the sensitivity of high quantiles from generalized Pareto distribution to changes in the threshold. As in the case of S&P500, for this analysis, we just only focus on the high quantiles (95%, 96%, 97%, 98% and 99%) as they are the only relevant quantiles in quantifying market risk. The quantile of the GPD is calculated using the expression (6). Figure 6 displays the GPD quantiles as a function of the threshold for all assets considered. Again, what pays our attention is that the line representing the quantiles as a function of the threshold is quite flat in the threshold range (80th to 96th). Just in the case of the high threshold, corresponding to the 97th to 99th percentiles, some differences are observed, especially for the threshold corresponding to the 99th percentile. For this threshold, the differences are around 25 basis points becoming 50 basis points for some assets as Merval index.

After checking that the GPD fits well the upper tail of the distribution of the assets for the set of thresholds considered, we calculate the market risk measure at the 95% and 99% confidence levels. For evaluating the accuracy of the VaR estimates, we use the standard tests that we presented in “[Backtesting](#)” section: LR_{uc} , LR_{ind} , LR_{cc} and DQ. For each asset, Table 8 displays the number of times that each of these tests is rejected for the 20 thresholds selected.

In the footnote in Table 8, we indicate the set of thresholds for which the null hypothesis is rejected. For instance, for CAC40 at a 95% confidence level, LR_{uc} test is rejected once for the threshold corresponding to the 99th percentile. The results obtained for VaR are as follows. According to LR_{uc} tests, in 7 of the 15 considered assets, we do not find evidence against the null hypothesis that the “VaR(5%) estimate is accurate”. This result is independent of the selected threshold, although, for certain indexes, this hypothesis is rejected for some tests performed over the threshold corresponding to the 99th percentile. The results found for VaR at the 99% confidence level are even more conclusive than those for VaR at the 95% confidence level. According to LR_{uc} tests, in 14 of the 15 considered assets, we do not find evidence against the null hypothesis that the “VaR (1%) estimate is accurate”. In certain cases, the accuracy tests provide evidence against the null hypothesis; however, in these cases, the rejection does not depend on the threshold selected. For instance, for the Cooper commodity, the DQ test rejects the null hypothesis for all thresholds. These results suggest that the quantification of the risk through the VaR measure does not depend on the threshold selected for this objective.

To test whether the ES estimations are correct, we use the procedure proposed by McNeil and Frey (2000) test. Overall, we do not find evidence against the null hypothesis that the average of the discrepancy measure is equal to zero indicating



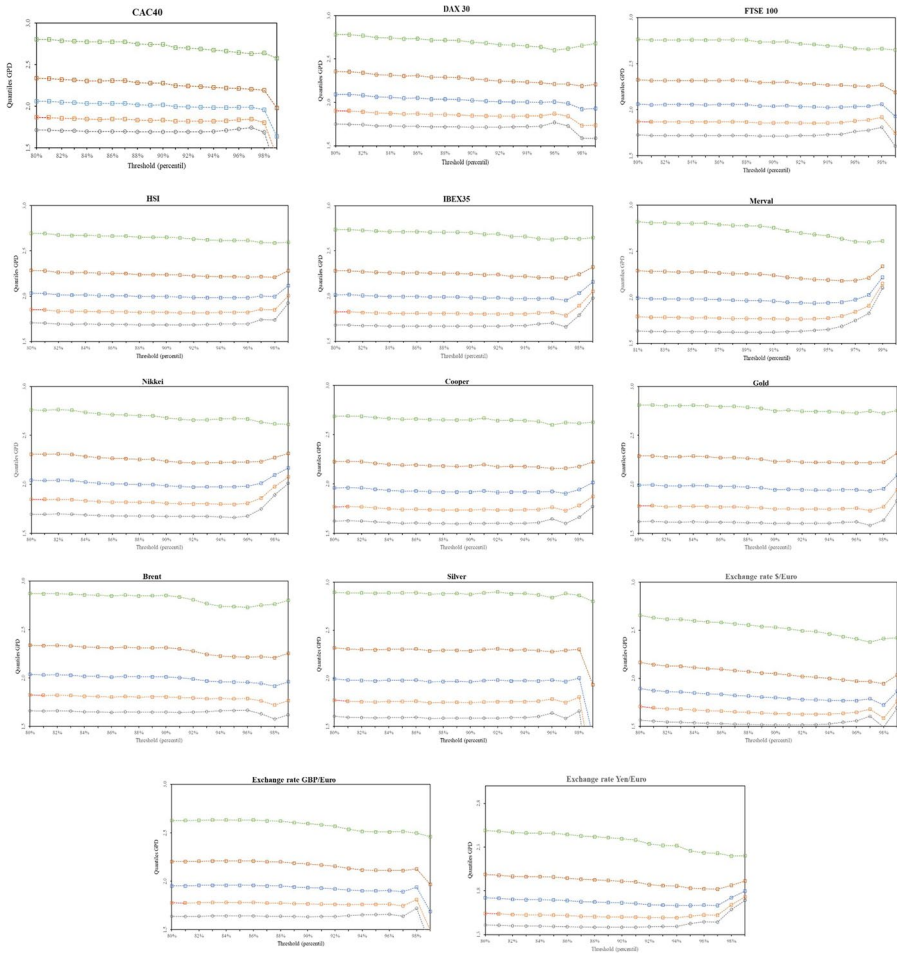


Fig. 6 GPD quantiles. GPD quantiles at 95% (black line) 96% (orange line) 97% (blue line) 98% (brown line) and 99% (green line) probability are displayed as a function of the thresholds

that all the threshold provide correct ES estimations for both 95% and 99% confidence level.

The results presented in this section corroborate those obtained for S&P500, indicating that the quantification of market risk through the VaR and ES measures does not depend on the threshold selected for applying the POT method.

To last, for all assets, we calculate the market risk capital requirement on the base of the ES at 99%. Table 9 shows the mean of these requirements. Again we find that in general for all assets there is a wide set of thresholds that give similar results. Just only the extreme thresholds provide capital requirements something different. If the aim of a financial institution is to minimize the market risk capital charges, the optimal threshold is the threshold corresponding to 90th percentile for six of the 14 assets considered. For three indexes (IBEX35, Merval and Nikkei), the highest



Table 8 Backtesting VaR and ES

	95% confidence level					99% confidence level				
	VaR				ES	VaR				ES
	LRuc	LRind	LRcc	DQ	MF	LRuc	LRind	LRcc	DQ	MF
CAC40	1 ⁽¹⁾	0	0	1 ⁽¹⁾	1 ⁽³⁾	0	0	0	0	0
DAX30	0	0	0	0	2 ⁽³⁾	0	0	0	0	0
FTSE100	0	0	0	0	19	0	0	0	0	0
HANG SENG	0	0	0	1 ⁽¹⁾	0	0	0	0	0	0
IBEX35	1 ⁽¹⁾	0	1 ⁽¹⁾	0	1 ⁽¹⁾	0	0	0	0	0
MERVAL	0	0	0	0	0	0	0	0	0	0
NIKKEI	2 ⁽²⁾	0	1 ⁽¹⁾	0	1 ⁽¹⁾	0	0	0	0	0
S&P500	0	0	0	20	15	0	0	0	0	0
COPPER	0	0	0	0	0	0	0	0	20	0
GOLD	1 ⁽¹⁾	0	0	0	0	0	0	0	0	0
OIL BRENT	2 ⁽²⁾	0	1 ⁽¹⁾	1 ⁽¹⁾	4 ⁽⁴⁾	17 ⁽⁴⁾	0	4 ⁽³⁾	0	0
SILVER	1 ⁽¹⁾	0	1 ⁽¹⁾	0	1 ⁽¹⁾	0	0	0	0	0
\$/€	1 ⁽¹⁾	0	1 ⁽¹⁾	0	0	0	0	0	0	0
£/€	2 ⁽²⁾	0	2 ⁽²⁾	0	0	0	0	0	2 ⁽²⁾	0
¥/€	0	0	0	0	0	0	0	0	1 ⁽¹⁾	0

The table counts the number of rejections for the 20 thresholds considered. Reject for: (1) threshold corresponding to 99th percentile ($\alpha=99\%$); (2) thresholds corresponding to 98th and 99th percentiles; (3) threshold corresponding to 97th percentile; (4) thresholds corresponding to 96th and 97th percentiles; (5) thresholds corresponding to the interval [84th, 98th]; (6) thresholds 81st and the interval [84th, 99th]; (7) thresholds in the percentiles range [95th, 98th]; (8) thresholds corresponding to the interval [93rd, 96th]; (9) All thresholds except for that corresponding to 99th percentile; (10) All thresholds. For the case of ES backtesting, the acceptance level of the null hypothesis is set at 1%

Table 9 Forecasting daily capital charges based on ES measure (99% confidence level)

Threshold	CAC40	DAX	FTSIE 100	HANG SENG	IBEX 35	MERVAL	NIKKEI	COPPER	GOLD	BRENT	SILVER	\$/€	£/€	¥/€
80 th	11.48	12.24	10.18	12.22	12.67	27.36	12.20	14.82	9.76	26.27	20.95	4.57	4.23	5.88
81 th	11.46	12.23	10.18	12.22	12.67	27.35	12.20	14.82	9.76	26.44	21.03	4.57	4.23	5.89
82 th	11.46	12.19	10.18	12.20	12.65	27.35	12.21	14.82	9.77	26.28	21.14	4.57	4.23	5.89
83 th	11.46	12.20	10.18	12.59	12.65	27.34	12.19	14.87	9.77	26.29	21.23	4.58	4.23	5.89
84 th	11.45	12.20	10.18	12.21	12.64	27.35	12.15	14.91	9.77	26.49	21.11	4.58	4.23	5.90
85 th	11.46	12.19	10.18	12.59	12.64	27.34	12.13	14.95	9.77	26.50	21.14	4.59	4.23	5.89
86 th	11.46	12.19	10.19	12.59	12.64	27.35	12.13	14.96	9.77	26.49	21.15	4.59	4.23	5.89
87 th	11.46	12.17	10.19	12.59	12.63	27.31	12.12	14.97	9.77	26.50	21.27	4.60	4.23	5.89
88 th	11.43	12.17	10.18	12.58	12.62	27.28	12.12	14.98	9.88	26.49	21.26	4.61	4.22	5.89
89 th	11.43	12.16	10.16	12.58	12.61	27.28	12.11	14.98	9.89	26.50	21.23	4.63	4.21	5.88
90 th	11.43	12.32	10.16	12.58	12.60	27.27	12.11	14.97	9.91	26.51	21.26	4.64	4.24	5.88
91 th	11.41	12.31	10.17	12.58	12.58	27.21	12.11	14.96	9.92	26.76	21.19	4.67	4.23	5.88
92 th	11.42	12.32	10.13	12.70	12.59	27.09	12.11	14.98	9.93	26.80	21.15	4.70	4.22	5.90
93 th	11.42	12.32	10.12	12.69	12.54	27.05	12.11	14.98	9.94	27.14	21.69	4.71	4.22	5.93
94 th	11.42	12.32	10.22	12.68	12.52	26.99	12.12	15.12	9.94	27.18	21.24	4.76	4.22	5.96
95 th	11.42	12.35	10.22	12.70	12.49	26.96	12.14	15.12	9.93	27.45	21.72	4.88	4.24	6.05
96 th	11.42	12.36	10.20	12.68	12.46	26.90	12.08	15.19	9.93	28.03	21.79	4.94	4.23	6.18
97 th	11.61	12.40	10.19	12.63	12.50	26.88	11.99	15.16	9.95	27.65	21.73	5.10	4.23	6.24
98 th	11.60	12.38	10.21	12.62	12.40	26.85	11.92	15.13	9.94	27.46	21.75	4.95	4.23	6.58
99 th	11.62	12.45	10.20	12.61	12.32	27.26	11.94	15.15	9.90	27.37	22.06	5.62	4.27	7.40

Note: The table shows the mean of forecasting DCR for all the assets considered. We remark in bold the minimum mean daily capital requirement and shadow the cases in which the differences regarding the 90th percentile are less than 5 basis points.



thresholds (98th and 99th) are the best, whilst for the commodities, the thresholds that minimise market risk capital requirement are the lowest(80th).

Conclusions

The conditional extreme value theory has been proven to be one of the most successful in estimating market risk. The implementation of this method in the framework of the POT model requires choosing a threshold return for fitting the generalized Pareto distribution. Threshold choice involves balancing bias and variance. To determine the optimal threshold, several techniques have been proposed such as graphic methods, ad hoc methods, or methods based on goodness-of-fit contrasts. However, none of these techniques have been proven to provide better results than others.

In this paper, we ask if the threshold choice is relevant in measuring market risk. In other words, in this study, we assess to what extent the selection of the threshold is decisive in quantifying the market risk. To measure market risk, we have used the value at risk (VaR) and expected shortfall (ES) measures. The study has been done for the S&P500index.

Previously, we analyse both, the sensitivity of the parameter estimates and GPD quantiles to the threshold choice. The results obtained are as follows. First, we find that following the literature, the parameter estimations are very sensitive to the selected threshold for estimating GPD. However, the quantiles of the GPD do not change much when the threshold changes, particularly for high quantiles (95th, 96th, 97th, 98th and 99th), which are relevant in risk estimation. Second, for a large set of thresholds (from the 80th percentile to the 96th percentile), the VaR estimations are practically equivalent. A similar finding occurs for the ES measure. In a last application, we calculate the market risk capital requirements on the base of the ES(99%) estimations. The results reveals that there is a set of thresholds which provides the same results finding some differences for the higher percentiles.

The results obtained indicate that from the market risk management point of view there is not an optimal threshold but that there is a set of optimal thresholds which provide similar market risk measures. Thus, we can conclude that in market risk estimation, the researchers and practitioners should not focus excessively on the threshold choice, as a wide range of them produce the same risk estimates.

To corroborate these results, we have extended the S&P500 index study to a set of 14 assets (stock market indexes, commodities and exchange rates). The results obtained for these assets corroborate the results obtained for S&P500.

To last, although overall the quantification of the risk does not depend on the threshold choice, for a certain threshold, some differences are found therefore, the financial institution may be interested in choosing the threshold that minimises the market risk capital requirement.

Acknowledgments The authors gratefully acknowledge financial support granted by the Ministerio de Ciencia e Innovación PID2020-113367RB-I00 and Young Talent Research Project UNED 2022 (076-044355 ENER-UK).



References

- Abad, P., and S. Benito. 2013. A detailed comparison of value at risk estimates. *Mathematics and Computers in Simulation* 94: 258–276. <https://doi.org/10.1016/j.matcom.2012.05.011>.
- Bader, B., J. Yan, and X. Zhang. 2018. Automated threshold selection for extreme value analysis via ordered goodness-of-fit tests with adjustment for false discovery rate. *The Annals of Applied Statistics* 12 (1): 310–329. <https://doi.org/10.1214/17-AOAS1092>.
- Balkema, A., and L. De Haan. 1974. Residual life time at great age. *The Annals of Probability* 2 (5): 792–804. <https://doi.org/10.1214/aop/1176996548>.
- Basel Committee on Banking Supervision. (1996). *Supervisory framework for the use of "Backtesting" in conjunction with the internal models approach to market risk capital requirements*. <https://www.bis.org/publ/bcbs22.pdf>
- Basel Committee on Banking Supervision. (2006). *International Convergence of Capital Measurement and Capital Standards A Revised Framework Comprehensive Version*. Available at: <https://www.bis.org/publ/bcbs128.pdf>
- Basel Committee on Banking Supervision. 2012. *Fundamental review of the trading book*. Basel: Bank for International Settlements (BIS).
- Basel Committee on Banking Supervision. (2016). *Minimum capital requirements for market risk*. Available at: <https://www.bis.org/bcbs/publ/d352.pdf>
- Basel Committee on Banking Supervision. 2017. *High-level summary of Basel III reforms*. Basel: Bank for International Settlements (BIS).
- Basel Committee on Banking Supervision. (2019). *Minimum capital requirements for market risk*. Available at: <https://www.bis.org/bcbs/publ/d457.pdf>
- Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels. 2004. *Statistics of extremes: Theory and applications*. London: Wiley.
- Beirlant, J., A. Kijko, T. Reynkens, and J.H.J. Einmahl. 2018. Estimating the maximum possible earthquake magnitude using extreme value methodology: The Groningen case. *Natural Hazards* 169: 1–23. <https://doi.org/10.1007/s11069-017-3162-2>.
- Bekiros, S.D., and D.A. Georgoutsos. 2005. Estimation of value-at-risk by extreme value and conventional methods: A comparative evaluation of their predictive performance. *Journal of International Financial Markets Institutions and Money* 15 (3): 209–228. <https://doi.org/10.1016/j.intfin.2004.05.002>.
- Benito, S., C. López-Martín, and M.Á. Navarro. 2017. The role of the skewed distributions in the framework of extreme value theory (EVT). *International Business Research* 10 (11): 88–102. <https://doi.org/10.5539/ibr.v10n11p88>.
- Black F. 1976. Studies in stock price volatility changes. Proceedings of the 1976 *Business Meeting of the Business and Economics Statistics Section*, American Association: 177–181.
- Caballero-Megido, C., J. Hillier, D. Wyncoll, L. Boshier, and B. Gouldby. 2018. Comparison of methods for threshold selection for extreme sea levels. *Journal of Flood Risk Management* 11 (2): 127–140. <https://doi.org/10.1111/jfr3.12296>.
- Carreau, J., P. Naveau, and L. Neppel. 2017. Partitioning into hazard subregions for regional peaks-over threshold modeling of heavy precipitation. *Water Resources Research, American Geophysical Union* 53 (5): 4407–4426.
- Cebrián, A.C., and J. Abaurrea. 2006. Drought analysis based on a marked cluster Poisson model. *Journal of Hydrometeorology* 7: 713–723.
- Cebrián, A.C., M. Denuit, and P. Lambert. 2003. Generalized Pareto fit to the society of actuaries large claims database. *North America Actuarial Journal* 7 (3): 18–36. <https://doi.org/10.1080/10920277.2003.10596098>.
- Chavez-Demoulin, V., and P. Embrechts. 2004. Smooth extremal models in finance and insurance. *Journal of Risk and Insurance* 71 (2): 183–199. <https://doi.org/10.1111/j.0022-4367.2004.00085.x>.
- Chang, C. L., Jiménez-Martín, J. Á., Maasoumi, E., McAleer, M., & Pérez-Amaral, T. (2019). Choosing expected shortfall over VaR in Basel III using stochastic dominance. *International Review of Economics & Finance*, 60, 95–113. <https://doi.org/10.1016/j.iref.2018.12.016>
- Cheng, L., and A. AghaKouchak. 2014. Nonstationary precipitation intensity-duration-frequency curves for infrastructure design in a changing climate. *Scientific Reports* 4: 7093. <https://doi.org/10.1038/srep07093>.



- Cheng, L., A. AghaKouchak, E. Gilleland, and R.W. Katz. 2014. Non-stationary extreme value analysis in a changing climate. *Climatic Change* 127 (2): 353. <https://doi.org/10.1007/s10584-014-1254-5>.
- Choulakian, V., and M.A. Stephens. 2001. Goodness-of-fit tests for the generalized Pareto distribution. *Technometrics* 43 (4): 478–484. <https://doi.org/10.1198/00401700152672573>.
- Christoffersen, P. 1998. Evaluating interval forecasting. *International Economic Review* 39: 841–862. <https://doi.org/10.2307/2527341>.
- Coles S. (2001). An introduction to statistical modeling of extreme values. *British library Cataloguing in Publication Data*, vol. 208, 78–84. *Springer Series in Statistics*.
- Cruz, M.G. 2002. *Modeling, measuring and hedging operational risk*, vol. 4. New York: Wiley.
- Cunnane, C. 1973. A particular comparison of annual maxima and partial duration series methods of flood frequency prediction. *Journal of Hydrology* 18 (3–4): 257–271. [https://doi.org/10.1016/0022-1694\(73\)90051-6](https://doi.org/10.1016/0022-1694(73)90051-6).
- Danielsson, J., L. de Haan, L. Peng, and C.G. de Vries. 2001. Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis* 76: 226–248. <https://doi.org/10.1006/jmva.2000.1903>.
- Davison, A., and R. Smith. 1990. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society* 52 (3): 393–442. <https://doi.org/10.1111/j.2517-6161.1990.tb01796.x>.
- De Zea Bermudez, P., M.A. Amaral Turkman, and K.F. Turkman. 2001. A predictive approach to tail probability estimation. *Extremes* 4: 295–314. <https://doi.org/10.1023/A:1016546027962>.
- Domonkos, P., and K. Piotrowicz. 1998. Winter temperature characteristics in central Europe. *International Journal of Climatology* 18: 1405–1417. [https://doi.org/10.1002/\(SICI\)1097-0088\(19981115\)18:13%3c1405::AID-JOC323%3e3.0.CO;2-D](https://doi.org/10.1002/(SICI)1097-0088(19981115)18:13%3c1405::AID-JOC323%3e3.0.CO;2-D).
- Drees, H., L. de Haan, and S. Resnick. 2000. How to make a Hill plot. *Annals of Statistics* 28 (1): 254–274. <https://doi.org/10.1214/aos/1016120372>.
- DuMouchel, M. 1983. Estimating the stable index α in order to measure tall thickness: A critique. *The Annals of Statistics* 11 (4): 1019–1031. <https://doi.org/10.1214/aos/1176346318>.
- Dupuis, D.J. 1999. Exceedances over high thresholds: A guide to threshold selection. *Extremes* 1 (3): 251–261. <https://doi.org/10.1023/A:1009914915709>.
- Embrechts, P., C. Klüppelberg, and T. Mikosch. 1997. Modelling extremal events. *British Actuarial Journal* 5 (2): 465–465. <https://doi.org/10.1017/S1357321700000532>.
- Engle, R., and S. Manganelli. 2004. CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics* 22 (4): 367–381. <https://doi.org/10.1198/073500104000000370>.
- Ergün, A.T., and J. Jun. 2010. Time-varying higher-order conditional moments and forecasting intraday VaR and expected shortfall. *The Quarterly Review of Economics and Finance* 50 (3): 264–272. <https://doi.org/10.1016/j.qref.2010.03.003>.
- Ferreira A., L. de Haan, and L. Peng. 2003. On optimising the estimation of high quantiles of a probability distribution. *Statistics: A Journal of Theoretical and Applied Statistics* 37 (5): 401–434. <https://doi.org/10.1080/0233188021000055345>.
- Fontnouvelle, P., E. Rosengren, and J. Jordan. 2007. Implications of alternative operational risk modeling techniques. NBER Chapters in: *The Risks of Financial Institutions*, 475–505. National Bureau of Economic Research, Inc.
- Gerstengarbe, F.W., and P.C. Werner. 1989. A method for statistical definition of extreme value regions and their application to meteorological time series. *Zeitschrift Fuer Meteorologie* 39 (4): 224–226.
- Gerstengarbe, F.W., and P.C. Werner. 1991. Some critical remarks on the use of extreme-value statistics in climatology. *Theoretical and Applied Climatology* 44 (1): 1–8. <https://doi.org/10.1007/BF00865546>.
- Hall, P. 1990. Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *Journal of Multivariate Analysis*. 32: 177–203. [https://doi.org/10.1016/0047-259X\(90\)90080-2](https://doi.org/10.1016/0047-259X(90)90080-2).
- Haigh, I., Nicholls, R., and Wells, N. (2010). Assessing changes in extreme sea levels: Application to the English Channel, 1900–2006. *Continental Shelf Research*, 30 (9), 1042–1055. <https://doi.org/10.1016/j.csr.2010.02.002>
- Ho, A.K., and A.T. Wan. 2002. Testing for covariance stationarity of stock returns in the presence of structural breaks: An intervention analysis. *Applied Economics Letters* 9 (7): 441–447. <https://doi.org/10.1080/13504850110090210>.



- Jobst, A.A. 2007. Operational risk—The sting is still in the tail but the poison depends on the dose. *Journal of Operational Risk* 2 (2): 3–59. <https://doi.org/10.5089/9781451868036.001>.
- Jorion, P. 2001. *Value at risk: The new benchmark for managing financial risk*. McGraw-Hill.
- Katz, R., M. Parlange, and P. Naveau. 2002. Statistics of extremes in hydrology. *Advances in Water Resources* 25 (8–12): 1287–1304. [https://doi.org/10.1016/S0309-1708\(02\)00056-8](https://doi.org/10.1016/S0309-1708(02)00056-8).
- Karpouzou, D.K., S. Kavalieratou, and C. Babajimopoulos. 2010. Trend analysis of precipitation data in Pieria region (Greece). *European Water* 30: 31–40.
- Kharin, V.V., F. Zwiers, X. Zhang, and M. Wehner. 2013. Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Climatic Change* 119: 345–357. <https://doi.org/10.1007/s10584-013-0705-8>.
- Kupiec, P. 1995. Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives* 2: 73–84. <https://doi.org/10.3905/jod.1995.407942>.
- Langousis, A., A. Mamalakis, M. Puliga, and R. Deidda. 2016. Threshold detection for the generalized Pareto distribution: Review of representative methods and application to the NOAA NCDC daily rainfall database. *Water Resources Research* 52: 2659–2681. <https://doi.org/10.1002/2015WR018502>.
- Lasch, P., M. Lindner, B. Ebert, M. Flechsig, F.W. Gerstengarbe, F. Suckow, and P.C. Werner. 1999. Regional impact analysis of climate change on natural and managed forests in the Federal State of Brandenburg, Germany. *Environmental Modeling and Assessment* 4 (4): 273–286. <https://doi.org/10.1023/A:1019024619886>.
- Li, F., P.H.A.J.M. van Gelder, R. Ranasinghe, D.P. Callaghan, and R.B. Jongejan. 2014. Probabilistic modelling of extreme storms along the Dutch coast. *Coastal Engineering* 86: 1–13. <https://doi.org/10.1111/jfr3.12296>.
- Loretan, M., and P.C.B. Phillips. 1994. Testing the covariance stationarity of heavy-tailed time series: An overview of the theory with applications to several financial datasets. *Journal of Empirical Finance* 1 (2): 211–248. [https://doi.org/10.1016/0927-5398\(94\)90004-3](https://doi.org/10.1016/0927-5398(94)90004-3).
- MacDonald, A., C.J. Scarrott, D. Lee, B. Darlow, M. Reale, and G. Russell. 2011. A flexible extreme value mixture model. *Computational Statistics & Data Analysis* 55 (6): 2137–2157. <https://doi.org/10.1016/j.csda.2011.01.005>.
- Madsen, H., C.P. Pearson, and D. Rosbjerg. 1997a. Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 2. Regional Modeling. *Water Resources Research* 33 (4): 759–769. <https://doi.org/10.1029/96WR03849>.
- Madsen, H., P.F. Rasmussen, and D. Rosbjerg. 1997b. Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 1. At-Site Modeling. *Water Resources Research* 33 (4): 747–757. <https://doi.org/10.1029/96WR03848>.
- McMillan, A., Batstone, C., Worth, D., Tawn, J., Horsburgh, K., & Lawless, M. (2011). Coastal flood boundary conditions for UK mainland and islands. Project SC060064/TR2: Design sea levels. <https://nora.nerc.ac.uk/id/eprint/19721>
- McNeil, A.J., and R. Frey. 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance* 7: 271–300. [https://doi.org/10.1016/S0927-5398\(00\)00012-8](https://doi.org/10.1016/S0927-5398(00)00012-8).
- McNeil, A.J., R. Frey, and P. Embrechts. 2005. Quantitative risk management: concepts, techniques, and tools. *Princeton Series in Finance*.
- Moscadelli, M. 2004. The modelling of operational risk: Experience with the analysis of the data collected by the Basel Committee. <https://doi.org/10.2139/ssrn.557214>
- Naveau, P., R. Huser, P. Ribereau, and A. Hannart. 2016. Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. *Water Resources Research* 52: 2753–2769. <https://doi.org/10.1002/2015WR018552>.
- Neftci, S. 2000. Value at risk calculations, extreme events and tail estimation. *The Journal of Derivatives*. 7 (3): 23–27. <https://doi.org/10.3905/jod.2000.319126>.
- Northrop, P.J., and C.L. Coleman. 2014. Improved threshold diagnostic plots for extreme value analyses. *Extremes* 17: 289–303. <https://doi.org/10.1007/s10687-014-0183-z>.
- Omran, M.F., and E. McKenzie. 2010. Heteroscedasticity in stock returns data revisited: Volume versus GARCH effects. *Applied Financial Economics* 10 (5): 553–560. <https://doi.org/10.1080/096031000416433>.
- Papalexiou, S.M., D. Koutsoyiannis, and C. Makropoulos. 2013. How extreme is extreme? An assessment of daily rainfall distribution tails. *Hydrology and Earth System Sciences* 17 (2): 851–862. <https://doi.org/10.5194/hess-17-851-2013>.



- Pickands, J. 1975. Statistical inference using extreme order statistics. *Annals of Statistics* 3: 119–131.
- Queensley C. Chukwudum, Peter Mwita & Joseph K. Mung'atu. 2019. Optimal threshold determination based on the mean excess plot. *Communications in Statistics—Theory and Methods*. <https://doi.org/10.1080/03610926.2019.1624772>.
- Reiss, R.D., and M. Thomas. 2007. *Statistical analysis of extreme values: With applications to insurance, finance, hydrology and other fields*. Boston: Birkhauser.
- Resnick, S.I. 1997. Heavy tail modeling and teletraffic data. *The Annals of Statistics* 25 (5): 1805–1869.
- Rodríguez, Gabriel. 2017. Extreme value theory: An application to the Peruvian stock market returns, *Revista de Métodos Cuantitativos para la Economía y la Empresa*, ISSN 1886-516X, Universidad Pablo de Olavide, Sevilla, vol. 23, 48–74. <http://hdl.handle.net/10419/195412>
- Ruggiero, P., P.D. Komar, and J.C. Allan. 2010. Increasing wave heights and extreme value projections: The wave climate of the US Pacific Northwest. *Coastal Engineering* 57 (5): 539–552. <https://doi.org/10.1016/j.coastaleng.2009.12.005>.
- Scarrot, A., and A. McDonald. 2012. A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT—Statistical Journal*. 10 (1): 33–60. <https://doi.org/10.57805/revstat.v10i1.110>.
- Schneider, L.F., A. Krajina, and T. Krivobokova. 2021. Threshold selection in univariate extreme value analysis. *Extremes* 24: 881–913. <https://doi.org/10.1007/s10687-021-00405-7>.
- Solari, S., and M.A. Losada. 2012. A unified statistical model for hydrological variables including the selection of threshold for the peak over threshold method. *Water Resources Research*. <https://doi.org/10.1029/2011WR011475>.
- Tanaka, S., and K. Takara. 2002. A study on threshold selection in POT analysis of extreme floods. *IAHS AISH Pub.* 271: 299–304.
- Thompson, P., Y. Cai, D. Reeve, and J. Stander. 2009. Automated threshold selection methods for extreme wave analysis. *Coastal Engineering* 56 (10): 1013–1021. <https://doi.org/10.1016/j.coastaleng.2009.06.003>.
- Wadsworth, J.L., and J.A. Tawn. 2012. Likelihood-based procedures for threshold diagnostics and uncertainty in extreme value modelling. *Journal of Royal Statistical Society. Series b. Statistical Methodology*. 74: 543–567. <https://doi.org/10.1111/j.1467-9868.2011.01017.x>.
- Wang, Q.J. 1991. The POT model described by the generalized Pareto distribution with Poisson arrival rate. *Journal of Hydrology* 129 (1–4): 263–280. [https://doi.org/10.1016/0022-1694\(91\)90054-L](https://doi.org/10.1016/0022-1694(91)90054-L).
- Wang, J. J., Jing, Y. Y., Zhang, C. F., & Zhao, J. H. (2009). Review on multi-criteria decision analysis aid in sustainable energy decision-making. *Renewable and sustainable energy reviews*, 13(9), 2263–2278. <https://doi.org/10.1016/j.rser.2009.06.021>
- Werner, P.C., and F.W. Gerstengarbe. 1997. Proposal for the development of climate scenarios. *Climate Research* 8: 171–182. <https://doi.org/10.3354/CR008171>.
- Wu, G. and Qiu, G. (2018). Threshold Selection for POT Framework in the Extreme Vehicle Loads Analysis Based on Multiple Criteria. *Shock and Vibration*, 2018, 4654659, 9 pages, 2018. <https://doi.org/10.1155/2018/4654659>
- Wyncoll, D., and B. Gouldby. 2013. Application of a multivariate extreme value approach to system flood risk analysis. *Journal of Flood Risk Management* 1: 3–10. <https://doi.org/10.1111/jfr3.12069>.
- Žiković, S., and B. Aktan. 2009. Global financial crisis and VaR performance in emerging markets: A case of EU candidate states-Turkey and Croatia. *Zbornik radova Ekonomskog fakulteta u Rijeci: časopis za ekonomsku teoriju i praksu* 27 (1): 149–170.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

