SORT 40 (2) July-December 2016, 303-320

Modelling extreme values by the residual coefficient of variation

Joan del Castillo[1,*] and Maria Padilla[2]

Abstract

The possibilities of the use of the coefficient of variation over a high threshold in tail modelling are discussed. The paper also considers multiple threshold tests for a generalized Pareto distribution, together with a threshold selection algorithm. One of the main contributions is to extend the methodology based on moments to all distributions, even without finite moments. These techniques are applied to euro/dollar daily exchange rates and to Danish fire insurance losses.

MSC: 62E20, 62E15, 62M02.

Keywords: Statistics of extremes, heavy tails, high quantile estimation, value at risk.

1. Introduction

Fisher and Tippett (1928) and Gnedenko (1943) show that, under regularity conditions, the limit distribution for the normalized maximum of a sequence of independent and identically distributed (iid) random variables (r.v.) is a member of the generalized extreme value (GEV) distribution with a cumulative distribution function

$$H_{\xi}(x) = \exp\{-(1+\xi x)^{-1/\xi}\}, \quad (1+\xi x) > 0,$$

where ξ is called *extreme value index*. This family of continuous distributions contains the Fréchet distribution ($\xi > 0$), the Weibull distribution ($\xi < 0$), and the Gumbell distribution ($\xi = 0$, as a limit case), see McNeil et al. (2005) and Gomes and Guillou (2015).

* *Corresponding author:* castillo@mat.uab.cat

[1] Department of Mathematics, Universitat Autònoma de Barcelona, 08193, Cerdanyola del Vallès (Barcelona), Spain.

[2] Department of Mathematics. Universitat Autònoma de Barcelona. 08193, Cerdanyola del Vallès (Barcelona). Spain.

Received: March 2016

Accepted: August 2016

The Pickands–Balkema–DeHaan theorem, see Embrechts et al. (1997) and McNeil et al. (2005), initiated a new way of studying extreme value theory via distributions above a threshold, which use more information than the maximum data grouped into blocks. This theorem is a very widely applicable result that essentially says that the generalized Pareto distribution (GPD) is the canonical distribution for modelling excess losses over high thresholds. The cumulative distribution function of GPD(ξ,ψ) is

$$F(x) = 1-(1+\xi x/\psi)^{-1/\xi}, \tag{1}$$

where $\psi > 0$ and ξ are scale and shape parameters. For $\xi > 0$ the range of x is $x > 0$, being in this case the usual Pareto distribution. The limit case $\xi = 0$ corresponds to the exponential distribution. For $\xi < 0$ the range of x is $0 < x < \psi/|\xi|$ and GPD has bounded support. The GPD has mean $\psi/(1-\xi)$ and variance $\psi^2/[(1-\xi)^2(1-2\xi)]$ provided $\xi < 1/2$.

Let X be a continuous non-negative r.v. with distribution function $F(x)$. For any threshold, $t > 0$, the r.v. of the conditional distribution of threshold excesses $X-t$ given $X > t$, denoted as $X_t = \{X-t \mid X > t\}$, is called the *residual distribution* of X over t. The cumulative distribution function of X_t, $F_t(x)$, is given by

$$1-F_t(x) = (1-F(x+t))/(1-F(t)). \tag{2}$$

The quantity $M(t) = E(X_t)$ is called the *residual mean* and $V(t) = \mathrm{Var}(X_t)$ the *residual variance*. The *residual coefficient of variation* (CV) is given by

$$\mathrm{CV}(t) \equiv \mathrm{CV}(X_t) = \sqrt{V(t)}/M(t), \tag{3}$$

like the usual CV, the function CV(t) is independent of scale, that is, if λ is a positive constant then $\mathrm{CV}(\lambda X_t) = \mathrm{CV}(X_t)$.

The residual distribution of a GPD is again GPD and for any threshold $t > 0$, the shape parameter ξ is invariant, in fact

$$\mathrm{GPD}_t(\xi,\psi) = \mathrm{GPD}(\xi,\psi+\xi t). \tag{4}$$

Note that the residual CV is independent of the threshold and the scale parameter, since it is given by

$$\mathrm{CV}(t) = c_\xi = \sqrt{1/(1-2\xi)}. \tag{5}$$

Gupta and Kirmani (2000) show that the residual CV characterizes the distribution in univariate and bivariate cases, provided that a finite second moment exists. In the case of GPD, the residual CV is constant and it is a one to one transformation of the extreme value index suggesting its use to estimate this index.

Castillo et al. (2014) suggest a new tool to identify the tail of a distribution based on the residual CV, henceforth called CV-plot, as an alternative to the *mean excess plot* (ME-plot), a commonly used diagnostic tool in risk analysis to justify fitting a GPD, see Ghosh and Resnick (2010), Embrechts et al. (1997) and Davison and Smith (1990). What is important here is the fact that for a GPD distribution with $\xi < 1$, the residual mean function $t \to M(t)$ is linear with positive, negative or zero slope depending on whether $0 < \xi < 1$, $\xi < 0$ or $\xi = 0$.

Given a sample $\{x_k\}$ of size n of positive numbers, we denote the ordered sample $\{x_{(k)}\}$, so that $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$. The CV-plot is the function $cv(t)$ of the sample coefficient of variation of the *threshold excesses* $(x_j - t)$ for the *exceedances* $\{x_j : x_j > t\}$ given by

$$t \to cv(t) = \frac{sd\{x_j - t \mid x_j > t\}}{\text{mean}\{x_j - t \mid x_j > t\}}, \tag{6}$$

in practice $t = x_{(k)}$ are the order statistics, where, k $(1 \leq k \leq n)$ is the size of the subsample removed. This tool has been applied to financial and environmental datasets, see Castillo and Serra (2015).

The CV-plot has two advantages over ME-plot: first, ME-plot depends on a scale parameter and CV-plot does not; second, linear functions are defined by two parameters and the constants by only one. So the uncertainty is reduced from three to one single parameter.

A unconscientious use of some measures of variation can lead to wrong conclusions, see Albrecher et al. (2010). A serious problem with the residual coefficient of variation is the fact that the proposed method only works when the extreme value index is smaller than 0.25 (otherwise its variance is not finite). To fix this, some transformations that relate light-heavy tails are introduced in Section 2.

Section 3 extends some results of Castillo et al. (2014) from the exponential distribution to all GPD when the extreme value index is below 0.25. Moreover, multiple threshold tests together with a threshold selection algorithm, designed in a way that avoids subjectivity, are also achieved. In Section 4, these techniques are applied first to euro/dollar daily exchange rates and validated with out of sample observations. Secondly, the approach developed in Section 2, is illustrated using the Danish fire insurance dataset, a highly heavy-tailed, infinite-variance model.

2. Transformations of heavy-light tails

The transformations introduced to this section make it possible to estimate the extreme value index using methods based on moments in situations where moments are not finite.

A distribution function F is said to be in the maximum domain of attraction of H_ξ, written $F \in D(H_\xi)$, if under appropriate normalization the block maxima of an iid sequence of r.v. with distribution F converge to H_ξ. For a r.v. X with distribution function F is also written $X \in D(H_\xi)$. A positive function L on $(0,\infty)$ *slowly varies* at ∞ if

$$\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1, \quad t > 0.$$

Regularly varying functions can be represented by power functions multiplied by slowly varying functions, i.e. $h(x) \in \mathrm{RV}_\rho$ if and only if $h(x) = x^\rho L(x)$.

Gnedenko proved, see McNeil et al. (2005, Theorems 7.8 and 7.10), that the maximum domain of attraction of a Fréchet distribution, with shape parameter $\xi > 0$, is characterized in terms of the tail function, $\overline{F}(x) = 1 - F(x)$, by

$$F \in D(H_\xi) \Leftrightarrow \overline{F} \in \mathrm{RV}_{-1/\xi} \quad (\xi > 0).$$

Similarly the maximum domain of attraction of a Weibull distribution, with shape parameter $\xi < 0$, is characterized by

$$F \in D(H_\xi) \Leftrightarrow \overline{F}(x_+ - 1/x) \in \mathrm{RV}_{1/\xi} \quad (\xi < 0),$$

where $x_+ = \sup\{x : F(x) < 1\}$.

The following result of practical importance is embedded in the previous characterizations, and which to our knowledge it has not been pointed out.

Corollary 1 *Let X be a continuous r.v. with cumulative distribution function F.*

(1) If $X \in D(H_\xi), X > 0$, with $\xi > 0$, then $X^* = -1/X \in D(H_{-\xi})$.

(2) If $X \in D(H_\xi)$ with $\xi < 0$, then $X^* = -1/(X - x_+) \in D(H_{-\xi})$, where $x_+ = \sup\{x : F(x) < 1\}$.

Proof. (1) The cumulative distribution function of X^* is $F^*(x) = F(-1/x)$ and $x_+ = \sup\{x : F^*(x) < 1\} = 0$. By assumption $\overline{F}(x) = x^{-1/\xi}L(x)$ with L slowly varying at ∞, hence $\overline{F^*}(x_+ - 1/x) = \overline{F}(x) = x^{1/(-\xi)}L(x)$ and $X^* \in D(H_{-\xi})$.

(2) Since the translation of a v.a. does not alter the domain of attraction, we can assume $x_+ = 0$ without loss of generality. The tail function of X^* is now $\overline{F^*}(x) = \overline{F}(-1/x) = x^{1/\xi}L(x)$. Hence, $\overline{F^*}(x) \in \mathrm{RV}_{1/\xi}$ and $X^* \in D(H_{-\xi})$. ∎

Corollary 1 provides an asymptotic tool and it is related to an exact result in the GEV model: X has Fréchet distribution if and only if $-1/X$ has Weibull distribution with the same extreme value index, but with the sign changed. However, the corresponding result is not true in GPD, as we discuss below.

For a r.v. X, the Pickands–Balkema–DeHaan theorem shows that $X \in D(H_\xi)$ if and only if the limiting behaviour of the residual distribution of X over t, X_t, is like a GPD with the same parameter ξ, see McNeil et al. (2005, Theorem 7.20). According to the Pickands–Balkema–DeHaan theorem, Corollary 1 can be interpreted as follows.

Corollary 2 *Let X be a continuous positive r.v. such that the limiting behaviour of the residual distribution of X over a threshold is* GPD *with parameter $\xi > 0$ ($\xi < 0$), then the limiting behaviour of the residual distribution of $-1/X$ ($-1/(X - x_+)$) over a threshold is* GPD *with parameter $-\xi$.*

Corollary 2 enables determination of the extreme value index for heavy tailed distributions using light tail models and vice versa. For instance ME-plot and CV-plot can be used to determine the extreme value index in really heavy tailed distributions, see the example 4.2 in Section 4. These asymptotic results can be improved on GPD for practical aplications.

The GPD(ξ, ψ) distributions are standardized so that all their observations take positive values. The supports of the distributions are $(0, \sigma)$, where $\sigma = \infty$ for $\xi \geq 0$ and $\sigma = \psi/|\xi|$ for $\xi < 0$. The GPD distributions can be extended to include a location parameter by $Y = X + \mu$. The behaviour of X near σ is the same as that of Y near $\sigma + \mu$. The transformation $X^* = -1/X$ is also associated with the origin at zero, but can be generalized to $Y = -1/(X + c)$, provided $c \geq 0$, or $c \leq -\sigma$, and the transformations is monotonous increasing on $(0, \sigma)$. The following result examines these transformations on GPD.

Theorem 3 *Let X be a r.v. with* GPD(ξ, ψ) *distribution in $(0, \sigma)$ and $c \geq 0$ or $c \leq -\sigma$, then $Y = -1/(X + c)$ has distribution* GPD *with location parameter if and only if $c = \psi/\xi$. Then $Z = Y + 1/c$ has* GPD$(-\xi, \xi^2/\psi)$ *distribution.*

Proof: From (1) the distribution function of Y is

$$F_Y(y) = F(x(y)) = 1 - \left(1 - \frac{\xi}{\psi}\left(\frac{cy+1}{y}\right)\right)^{-1/\xi} = 1 - \left(\frac{\psi y}{y(\psi - \xi c) - \xi}\right)^{1/\xi}, \tag{7}$$

where $-1/c < y < -1/(\sigma + c)$. The denominator of the right term of (7) is a constant if and only if $c = \psi/\xi$. In this case the distribution function of Z is

$$F_Z(z) = F_Y(y(z)) = 1 - (1 - \psi z/\xi)^{1/\xi} = 1 - (1 - \xi z/(\xi^2/\psi))^{1/\xi},$$

where $0 < z < \sigma_z$, $\sigma_z = \xi/\psi$ for $\xi > 0$ and $\sigma_z = \infty$ for $\xi < 0$. Hence, Z has GPD$(-\xi, \xi^2/\psi)$ distribution and Y has GPD distribution with location parameter $-1/c$. ■

The following result establishes the essential content of the Theorem 3 avoiding the location parameter.

Corollary 4 *Let $\xi > 0$, $\psi > 0$ and $c = \psi/\xi$, then a r.v. X has* GPD(ξ,ψ) *distribution if and only if $Z = X/(c(X+c))$ has* GPD(ξ_z,ψ_z) *distribution with $\xi_z = -\xi$, $\psi_z = \xi^2/\psi$ and the support $(0,\xi/\psi)$.*

Proof: In one sense, this is proved by Theorem 3, because $c > 0$ and $Z = X/(c(X+c)) = -1/(X+c)+1/c$.

The converse is also a consequence of Theorem 3, because the inverse of the above transformation is

$$X = c^2Z/(1-cZ) = Z/(c_2(Z+c_2)) = -1/(Z+c_2)+1/c_2$$

where $c_2 = -1/c = -\xi/\psi$. The support of Z is $(0,\psi_z/|\xi_z|) = (0,\xi/\psi)$ and $Z+c_2 < 0$ (equivalently $c_2 \le -\xi/\psi$), then X is a monotonous increasing function of Z and Theorem 3 proves the result. ■

3. Multiple threshold test

In this Section, the asymptotic distribution of the residual coefficient of variation for GPD is studied as a random process indexed by the threshold. This provides pointwise error limits for CV-plot and the multiple thresholds test that really reduce the multiple testing problem, hence, the p-values are clearly defined. These results generalize and summarize some of those of Castillo et al. (2014) on the the exponential distribution. Multiple test are often used on testing extreme value copulas, see Bahraoui et al. (2014).

Theorem 5 *Let $\{X_j\}$ be a sample of size n of iid* GPD(ξ,ψ) *distributed r.v., with $\xi < 1/4$. Then $\sqrt{n}(cv(t)-c_\xi)$, where $cv(t)$ and c_ξ were respectively defined in* (6) *and* (5)*, converges in finite-dimensional distributions to a Gaussian process with zero mean and covariance function given by*

$$\rho_0(s,t) = \exp(\min(s,t)/\psi),$$

for $\xi = 0$, and

$$\begin{aligned}\rho_\xi(s,t) = {} & (((\psi+\xi s)/\psi)^{1/\xi})(1-\xi)^2(6\xi^4t^2+12\psi\xi^3t+8\xi^3st-9\xi^3t^2+6\psi^2\xi^2 \\ & +8\psi\xi^2s-10\psi\xi^2t-2\xi^2st+3\xi^2t^2-\psi^2\xi-2\psi\xi s+4\psi\xi t+\psi^2) \\ & /((1-3\xi)(1-2\xi)^2(1-4\xi)(\psi+\xi s)^2)\end{aligned}$$

for $\xi \neq 0$ and $s \le t$.

Proof: See Appendix A. ■

Pointwise error limits of the CV-plot under GPD follow from the next result.

Corollary 6 *Given a sample* $\{X_j\}$ *of a* GPD(ξ,ψ) *distribution* $(\xi < 1/4)$ *and a fixed threshold* t*, the asymptotic distribution of the residual* CV *is*

$$\sqrt{n(t)}(cv(t) - c_\xi) \xrightarrow{d} N(0, \sigma_\xi^2). \tag{8}$$

where c_ξ *is in* (5)*,* $n(t) = \sum_{j=1}^{n} 1_{(X_j > t)}$ *and*

$$\sigma_\xi^2 = \frac{(1-\xi)^2(6\xi^2 - \xi + 1)}{(1-2\xi)^2(1-3\xi)(1-4\xi)}.$$

Proof: The proof follows directly from Corollary 2 in Castillo et al. (2014). The asymptotic variance is given by $\sigma_\xi^2 = \rho_\xi(0,0)$, where the covariance function is in Theorem 5. The Theorem 5 can be applied to the threshold excesses $\{X_j - t \mid X > t\}$, replacing n with $n(t)$ and $cv(0)$ with $cv(t)$. From (4) the threshold excesses are again GPD with the same parameter ξ and the CV does not depend on ψ. ∎

From the last result the asymptotic confidence intervals of the CV-plot for exponential distribution are obtained taking $c_0 = 1$ and $\sigma_0^2 = 1$ and for uniform distribution taking $c_{-1} = 1/\sqrt{3}$ and $\sigma_{-1}^2 = 8/45$.

Corollary 6 needs a fixed value ξ and a fixed threshold t. However, in order to have a consistent test in GPD, $\mathrm{CV}(t) = c_\xi$ must be checked for all of threshold t, in accordance with the characterization by Gupta and Kirmani (2000). For instance, the absolute value of the Student t_4 distribution has CV equal to 1 and can not be distinguished from the exponential distribution with a direct application of Corollary 6.

3.1. Exact null hypothesis test

In order to test whether a sample $\{x_j\}$ of size n of non-negative numbers, is distributed as a GPD with parameter ξ, a set of thresholds $th = \{0 = t_0 < t_1 < \cdots < t_m\}$ will be selected to test the null hypothesis

$$H_0 : \mathrm{CV}(t_k) = c_\xi, \quad k = 0, 1, \ldots, m.$$

Hence, if H_0 is accepted and m is large enough, say 20 or 50, it will be more reasonable to assume that the sample comes from a distribution GPD(ξ,ψ) than from applying Corollary 6 to a single threshold.

Let us denote $D_t(\xi) \equiv \sqrt{n(t)}(cv(t) - c_\xi)$, from Corollary 6, $D_t^2(\xi)/\sigma_\xi^2$ has asymptotic distribution χ_1^2 under the null hypothesis of GPD $(\xi < 0.25)$. Let us denote

$$T_{th}(\xi) = \sum_{k=0}^{m} D_{t_k}^2.$$

The distribution of $T_{th}(\xi)$ is independent from the scale parameter ψ under the null hypothesis of GPD. Then, its asymptotic expectation is $(m+1)\sigma_\xi^2$ and $T_{th}(\xi)/(m+1)$ is an estimator of the asymptotic variance σ_ξ^2, when ξ is known or estimated.

Given a sample $\{x_j\}$ of size n of non-negative numbers, $Q_n(p)$ denotes the inverse of the empirical distribution function,

$$Q_n(p) = \inf[x : F_n(x) \geq p]. \tag{9}$$

From a set of probabilities $\{0 = p_0 < p_1 < \cdots < p_m\}$ let $qu = \{0 = q_0 < q_1 < \cdots < q_m\}$ be the corresponding empirical quantiles of the sample, $q_k = Q_n(p_k)$, that will be used like the previous thresholds. Let us denote

$$T_{qu}(\xi) = \sum_{k=0}^{m} D_{q_k}^2.$$

$T_{qu}(\xi)$ is a multiple thresholds invariant statistic when the sample is multiplied by a positive number while maintaining the set of probabilities, since the empirical CV is invariant. This first condition ensures that the test results do not depend on units used for the observations.

A second desirable condition is to select the set of probabilities that determine the statistic $T_{qu}(\xi)$ so that the corresponding thresholds are approximately equally spaced. This can be achieved for the exponential distribution by taking $0 < p < 1$, $p_k = 1 - p^k$, $(k = 0, \ldots, m)$ and q_k as the corresponding quantiles. Since for a random variable X, distributed as an exponential with expected value μ, its quantile function is $Q(p) = \mu \log(1/p)$ and $\Pr\{X > (\mu \log(1/p))k\} = p^k$. Selecting the probabilities this way, $q_k = Q_n(p_k) \approx x_{(n-np^k)}$, $n(q_k) \approx n\,p^k$ and $T_{qu}(\xi)$ becomes

$$T_m(\xi) = n \sum_{k=0}^{m} p^k (cv(q_k) - c_\xi)^2. \tag{10}$$

In applications, given the number of single tests that will be included in the multivariant test, m, we choose the value of p, which determines the distance between the quantiles, such that $n\,p^m \approx n_s$, where n_s is the sample size such that for smaller subsamples CV is not accurate enough. Hence, given m, $p = (n_s/n)^{1/m}$ is suggested. In this paper $n_s \approx 8$ is used in numerical algorithms. Note that this way $T_m(\xi)$ depends only on ξ and m and the researcher chose only the number of thresholds used in the analysis, essentially eliminating subjectivity. These multiple thresholds tests generalize those developed by Castillo et al. (2014) for $\xi = 0$ and $p = 1/2$.

The asymptotic distribution of $T_m(\xi)$ is easily calculated from Theorem 5, following the steps suggested by Castillo et al. (2014), whenever $\xi < 0.25$. However, taking into account the different values of the extreme value index and the diverse small sample sizes, it is easier in practice to calculate the p-value for $T_m(\xi)$ using simulation methods, which are especially simple in this case. Assuming GPD for simulations, only the sample size, the number of thresholds, m, and ξ are needed. Since the distribution does not depend on scale, parameter $\psi = 1$ will be used.

3.2. Composite null hypothesis test

In most cases the parameter ξ is unknown and its estimate should be incorporated in the statistic $T_m(\xi)$ (see the R code in Appendix B). The method for estimating ξ leads to slight variations in the statistic, leading to essentially equivalent inference whenever we use the same estimation method in simulations to obtain the p-value. The null hypothesis is now that the sample comes from a distribution in which all $(m+1)$ residual CV are equal.

$$H_0 : \mathrm{CV}(q_0) = \cdots = \mathrm{CV}(q_m), \quad k = 0, 1, \ldots, m.$$

The alternative hypothesis is that the residual CV are equal from a threshold q_r $(0 < r \leq m)$ to the threshold q_m.

The most recommended estimation method is maximum likelihood estimation (MLE), although in GPD it is only asymptotically efficient provided $-0.5 < \xi$, see Davison and Smith (1990). For this distribution, the CV is a one-to-one transformation of ξ, see (5), and the empirical CV of the residual sample, $\mathrm{CV}(t)$, provides an alternative method of estimation. It is asymptotically normal whenever $\xi < 0.25$, see Corollary 6. The multiple thresholds tests (10) suggest estimating ξ as the value such that c_ξ achieves the minimum $T_m(\xi)$, namely

$$\tilde{c}_\xi = \sum_{k=0}^{m} p^k cv(q_k) / \sum_{k=0}^{m} p^k = (1-p) \sum_{k=0}^{m} p^k cv(q_k)/(1-p^{m+1}), \tag{11}$$

and reversing (5) provides $\tilde{\xi}$; standard errors of this estimator are readily provided by simulation. The main advantage of this method is that under the alternative hypothesis it is a better estimator than CV or MLE, since the sample is only GPD over a threshold q_r. Since the main interest is in samples that are not GPD, but in the tail, and results are often used in small samples with $\xi < 0$, the estimation method (11) is included in (10). Hence, the statistics for composite null hypothesis, that only depends on m, is $T_m = T_m(\tilde{\xi})$ given by

$$T_m = n \sum_{k=0}^{m} p^k (cv(q_k) - \tilde{c}_\xi)^2. \tag{12}$$

The R code for T_m used in the algorithms is in Appendix B.

3.3. Threshold Selection Algorithms

To select the number of extremes used in applying the peaks over a high threshold method, threshold selection algorithms are developed in this section to estimate the point above which the GPD distribution can be used to estimate the extreme value index for a set of extreme events, $\{x_j\}$, of size n. For this purpose the previous statistical tests will be adapted.

Note that in the T_m calculation the number of thresholds m is the only parameter that must be fixed by the researcher. This determines the thresholds (quantiles) where the CV is calculated, $\{0 = q_0 < q_1 < \cdots < q_m\}$, which are fixed throughout the procedure. Then, by simulation of GPD, the associated p-value is calculated (running 10^4 samples). After that, we accept or reject the null hypothesis with the estimated shape parameter using all the thresholds.

If the hypothesis is rejected, the threshold excesses $\{x_j - q_1\}$ are calculated for the sub-sample $\{x_j \geq q_1\}$. The previous steps are repeated, but removing one threshold, to accept or reject the null hypothesis that the sample comes from a GPD. At every stage only statistics associated to thresholds $k = r, \ldots, m$, where $0 \leq r \leq m$, are calculated:

$$T_m^r(\tilde{\xi}) = n \sum_{k=r}^{m} p^k (cv(q_k) - \tilde{c}_\xi)^2. \tag{13}$$

In summary, the **steps of the general algorithm** are

(1) Given m find p such that $np^m \approx n_s$, where n_s is the smaller sample size used to calculate CV (here $n_s = 8$ is used, but it can be modified).

(2) Calculate $\{0 = p_0 < p_1 < \cdots < p_m\}$, where $p_k = 1 - p^k$, and $\{0 = q_0 < q_1 < \cdots < q_m\}$, where $q_k = Q_n(1 - p^k)$, $k = 1, \ldots, m$.

(3) Estimate $\tilde{\xi}$ minimizing the value of $T_m(\xi)$ with the specific values in the previous steps.

(4) Calculate by simulation of GPD the p-value associated to the minimum $T_m(\tilde{\xi})$ and accept or reject the null hypothesis with the estimated shape parameter using all the thresholds (starting with $q_0 = 0$).

(5) If the hypothesis is rejected, compute the threshold excesses $\{x_j - q_1\}$ for the sub-sample $\{x_j \geq q_1\}$ and repeat the previous steps with $\{p_1 < \cdots < p_m\}$ and $\{q_1 < \cdots < q_m\}$, to accept or reject the null hypothesis that the sample comes from a GPD, but removing a threshold.

(6) Continue the process for the next value in the index of thresholds while the hypothesis is rejected.

Several authors recommend giving a prominent role to the exponential distribution in the model GPD, see Castillo and Serra (2015). The usual method for doing this is to consider the exponential models as the null hypothesis testing against GPD, see Kozubowski et al. (2009). Alternatively, one can consider the Akaike or Bayesian information criteria for model selection, see Clauset et al. (2009). The previous algorithm can be adapted to the case when $\xi = 0$ (or any known parameter) skipping step-3.

4. Fitting GPD to empirical data

In this Section, the methods developed previously are applied to two classic examples. The first one, the euro/dollar daily exchange rates between 1999 and 2005, is analyzed in the literature using distributions with heavy tails, when these models are not appropriate. Our methodology clearly shows this fact, see Figure 1. In addition, the analysis is validated with *out of sample* observations between 2005 and 2014, including the financial crisis of 2007-08.

For the second example, the Danish fire insurance dataset, the fitted model is a highly heavy-tailed, infinite-variance model. Hence, the methodology developed in Section 2 is needed to avoid unconscientious use of measures of variation that can lead to wrong connclusions Albrecher et al. (2010).

4.1. EUR/USD daily exchange rates

Gomes and Pestana (2007), introduce a new semi-parametric quantile estimation method based on an adequate bias-corrected Hill estimator. To illustrate their method it is applied to the analysis of log-returns of the euro/dollar (EUSD) daily exchange rates, from January 4, 1999 through November 17, 2005 (1,794 observations). The paper gives the estimations of the tail index $\hat{\xi} = 0.279$ (Hill estimator) and $\hat{\xi} = 0.247$ (bias-corrected) for the positive log returns of EUSD.

It should be mentioned that the Hill method always provide estimators with $\xi > 0$, as in this case. Hence, previously, this hypothesis has to be checked. Figure 1 shows the CV-plots (6) for the positive and negative (with the sign changed) log-returns of EUSD. In both cases there is empirical evidence that the residual CV is lower than 1. Since in GPD CV < 1 is equivalent to $\xi < 0$, this suggests light tails where some researchers assume heavy tails. This qualitative approach can be confirmed with the multiple thresholds tests.

Applying T_m, where $m = 20$, to the 900 positive log-returns of EUSD, the estimate of CV given by (11) is $\tilde{c}_\xi = 0.861$, which corresponds to $\tilde{\xi} = -0.174\ (0.031)$ assuming GPD. The statistic is $T_m = 6.435$ with a p-value of 0.421. Hence, the null hypothesis of GPD is not rejected for the entire sample and the previous estimation of ξ is validated (in the first step of the algorithm). The result is similar for the 874 negative log-returns and $m = 20$. Here $\tilde{c}_\xi = 0.868$ is obtained, which corresponds to $\tilde{\xi} = -0.163\ (0.032)$

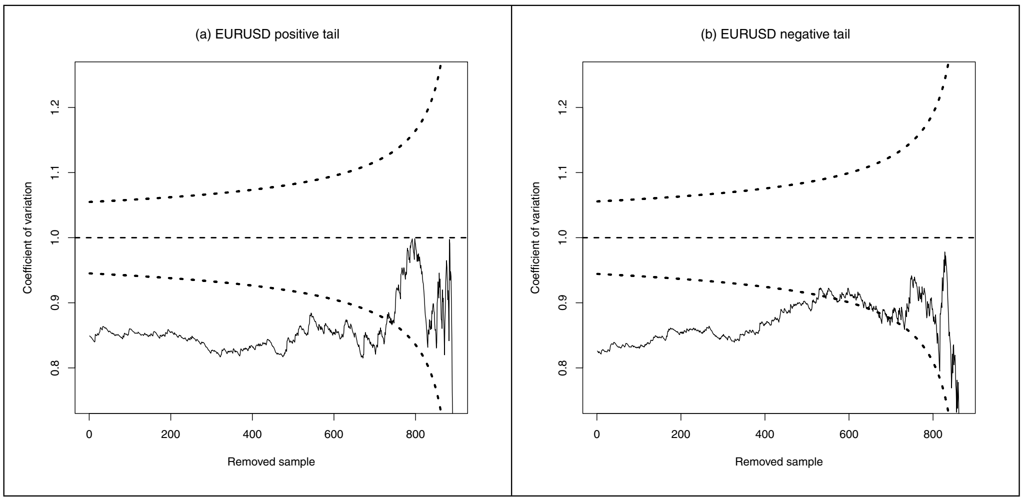

***Figure 1:** Residual empirical* CV *for positive and negative tails of EUR/USD daily exchange rates from January 4, 1999 through November 17, 2005. The dotted lines correspond to the asymptotic confidence intervals (*90%*) under exponentially* (CV = 1).

assuming GPD. The statistic is $T_m = 6.120$ with a p-value of 0.449. The null hypothesis of GPD is not rejected for the entire sample and the previous estimation of ξ is also validated. The results are fairly coincidental for $m = 10$, 20, 30 and 50, in both cases.

Despite the evidence of light tails on this previous sample, it is better to follow the recommendation of testing exponentiality at the tails. This approach is also compatible with the assumption of heavy tails in a wider sense ($\xi \geq 0$) including the exponential as a boundary point, see Castillo and Serra (2014). Applying T_m to all positive log-returns of EUSD, with $m = 20$ and $\xi = 0$, the null hypothesis of exponentiality is rejected (p-value 0.01). Taking $p = (n_s/n)^{1/m}$ the sample is reduced by $(1-p) = 21\%$ in each step, then for thresholds 0.134, 0.249 and 0.376, the null hypothesis is also rejected (p-values 0.017,0.026, and 0.057). Finally, exponentiality is not rejected over the threshold $t_p = 0.516$ with a p-value 0.133. For negative log-returns with $m = 20$ and $\xi = 0$, the exponentiality is rejected in the first three steps and not rejected over $t_n = 0.411$ with a p-value 0.126.

The main objective of statistics of extremes lies in the estimation of quantities related to extreme events that may happen in the future. Hence, the real challenge is to compare the results in out of sample observations. To this end, from the previous analysis, the *value at risk* at a level α (VaR_α), the quantile so that the chance of exceedance of that value is equal to α, is estimated by the *peak-over-threshold* method, using the empirical sample in the interval $(0,t)$, up to the estimated threshold, and the exponential distribution over threshold t. For $\alpha = 0.05$, 0.01 and 0.001, the quantiles of positive log-returns of EUSD are 1.316, 1.937 and 2.824; for the negative log-returns they are 1.352, 2.010 and 2.950.

Then, daily exchange rates, from November 18, 2005 through January, 14, 2014 (2,128 observations), including the financial crisis of 2007-08, are used as out of sample observations to assessing the predictive ability of the estimation of quantiles under the first dataset.

Using these 2,128 out of sample observations (the second dataset), the number of empirical exceedances of the last VaR_α estimations (under the first dataset, at 5%, 1%, 0.1%) are 42, 13, and 2, for the 1,080 positive log-returns (expected values 54.0, 10.8 and 1.1); and 47, 11 and 0, for the 1023 negative log-returns (expected values 51.2, 10.2 and 1.0). These results are fairly satisfactory and it can be concluded that the EUR / USD exchange has daily log-returns with exponential tails, including the financial crisis of 2007-08.

4.2. Danish fire insurance data

An interesting aspect of this article is the combination of the results of sections 2 and 3 when applying the peaks over threshold technique for tails in any maximum domain of attraction, even without finite moments. This approach is illustrated here using a classical example analyzed in several books and articles.

The Danish fire insurance data are a well-studied set of losses to illustrate the basic ideas of extreme value theory. The dataset consists of $2,156$ fire insurance losses over one million Danish kroner from 1980 to 1990 inclusive, see Embrechts et al. (1997, Example 6.2.9), Resnick (1997) and McNeil et al. (2005, Example 7.23).

In this example the authors agree to assume iid observations and a heavy tailed model. They also agree to set the threshold at $t = 10$ million Danish kroner, the exceedances over the threshold, denoted $\{x_j\}$, are $n_{10} = 109$. Fitting a GPD to $\{x_j\}$ by MLE, the parameter estimates in McNeil et al. (2005) are $\hat{\xi} = 0.50$ and $\hat{\psi} = 7.0$ with standard errors 0.14 and 1.1, respectively. Thus the fitted model is a very heavy-tailed, infinite-variance model and the methods in Section 3 cannot be applied directly. However, they can be used through the results shown in Section 2.

First of all, let us suppose we want to use CV to check whether the above data correspond to a GPD distribution with the estimated extreme value index. Applying Theorem 3 with $c = \hat{\psi}/\hat{\xi} = 14$, let $z_j = -1/(x_j + c) + 1/c$ be, then the set $\{z_j\}$ has light tails and the same extreme value index with the sign changed, provided that the estimated parameters are the true parameters. The CV of $\{z_j\}$ is $cv = 0.697$ which provides a new estimation of ξ, solving (5) by $\xi_z = (cv^2 - 1)/(2cv^2) = -0.530$, then, according to Theorem 3, $\tilde{\xi} = -\xi_z = 0.53$, not far from the parameter estimation in McNeil et al. (2005) , 0.50, since his standard error was 0.14. Alternatively, the multiple thresholds statistic T_m, from (13), can be used to check $\xi = 0.5$. The corresponding CV under GPD is $c_\xi = 0.707$. Taking $m = 20$, we get $T_m = 4.89$ with a p-value 0.421 (by simulation with 10^4 samples), not rejecting the null hypothesis.

Now consider the problem of choosing the threshold to estimate the extreme value index. In this example, most researchers use a visual observation of the ME-plot on

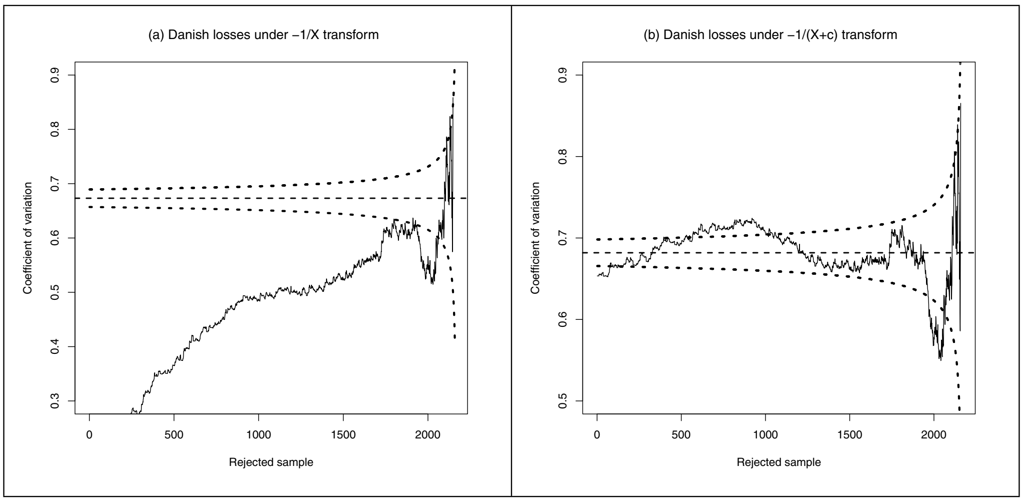

Figure 2: *Residual empirical* CV *for The Danish fire insurance losses under transformation of the data. (a): Dataset, transformed by* $X^* = -1/X$ *. (b): Dataset, transformed by* $Z = -1/(X + \psi/\xi)$*. The dotted lines correspond to the asymptotic confidence intervals* (90%) *under the estimated parameter, the dashed line is its* CV.

the full Danish dataset. The algorithm in Section 3.3 with the transformations from Section 2, comes to similar solutions automatically and opens up new perspectives.

Figure 2 shows the CV-plots of the full Danish dataset, transformed according to the Corollary 2, plot (a), and Theorem 3, plot (b). The first, corresponding to the transformation $X^* = -1/X$, shows an increasing CV and the second, corresponding to $Z = -1/(X+c)+1/c$, shows a stabilized CV close to a constant, indicating that the original dataset is close to a GPD, which is also shown by ME-plot.

Applying the algorithm of Section 3.3 with $m = 20$ after transformation X^*, constant residual CV is rejected in the first 11 steps (each one reduces the sample size by $(1-p) = 24\%$). Step 12, for the last 106 observations, accepts constant residual CV (p-value $= 0.269$) with estimates $\tilde{c}_\xi = 0.673$ and $\tilde{\xi} = 0.603$. The estimated threshold is approximately the same ($t = 10.2$ instead of 10), while the extreme value index is different but within the confidence interval.

The algorithm in Section 3.3, with $m = 20$ after transformation Z with $c = = 0.932/0.611 = 1.524$, rejects constant residual CV in the first three steps. Step 4, for the last 951 observations, accepts constant residual CV (p-value $= 0.167$) with estimates $\tilde{c}_\xi = 0.675$ and $\tilde{\xi} = 0.599$. The number of observations is much higher, the extreme value index being very close to that obtained with the transformation X^* and within the confidence interval. The p-value remains similar in the following steps up until the 12th, where it jumps up to 0.474. The number of observations is again 106 and the estimation $\tilde{\xi} = 0.548$, close to 0.50.

The conclusions from using the new methodology to analyze this dataset are the following. First, the results obtained by previous investigators are validated, in particular

GPD can be accepted with parameter $\xi = 0.5$, for the 109 larger observations see McNeil et al. (2005). This also shows the consistency of the presented methodology with other common techniques.

Moreover, from examining the extreme value index it is now known that for the 951 larger observations GPD can also be accepted, where the MLE parameter estimate is $\hat{\xi} = 0.680$, with standard error 0.055 ($\tilde{\xi} = 0.599$ obtained by T_m is within the confidence interval). The estimated extreme value index is now much more accurate because the sample size is much larger. We also note that the tails are heavier than was assumed, which means that higher risks should be considered.

Acknowledgement

This work was supported by the Spanish Ministry of Economy and Competitiveness under Grants: Applied Stochastic Processes, MTM 2012-31118 and Statistical modelling of environmental, technological and health risks, MTM2015-69493-R.

References

Albrecher, H., Ladoucette, S. and Teugels, J. (2010). Asymptotics of the sample coefficient of variation and the sample dispersion. *Journal of Statistical Planning and Inference*, 140, 358–368.

Bahraoui, Z., Bolance, C. and Perez-Marin, A. (2014). Testing extreme value copulas to estimate the quantile. *SORT*, 38, 89–102.

Castillo, J. D., Daoudi, J. and Lockhart, R. (2014). Methods to distinguish between polynomial and exponential tails. *Scandinavian Journal of Statistics*, 41, 382–393.

Castillo, J. D. and Serra, I. (2015). Likelihood inference for generalized Pareto distribution. *Computational Statistics and Data Analysis*, 83, 116–128.

Clauset, A., Shalizi C. and Newman, M. (2009). Power-law distributions in empirical data. *SIAM Review*, 51, 661–703.

Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B*, 393–442.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin.

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In: *Mathematical Proceedings of the Cambridge Philosophical Society*, 24, 180–190, Cambridge University Press.

Ghosh, S. and Resnick, S. (2010). A discussion on mean excess plot. *Stochastic Processes and their Applications*, 120, 1492–1517.

Gilleland, E., Ribatet, M. and Stephenson, A. G. (2013). A software review for extreme value analysis. *Extremes*, 16, 103–119.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *The Annals of Mathematics*, 44, 423–453.

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

Gomes, M. I. and Pestana, D. (2007). A sturdy reduced-bias extreme quantile (VaR) estimator. *Journal of the American Statistical Association*, 102, 280–292.

Gupta, R. C. and Kirmani, S. N. U. A. (2000). Residual coefficient of variation and some characterization results. *Journal of Statistical Planning and Inference*, 91, 23–31.

Kozubowski, T. J., Panorska, A. K., Qeadan, F., Gershunov, A. and Rominger, D. (2009). Testing exponentiality versus Pareto distribution via likelihood ratio. *Communications in Statistics-Simulation and Computation* 38, 118–139.

McNeil, A. J., Frey, R. and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton Series in Finance, New Jersey (2005).

R Core Team (2010). *R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing*. Vienna, Austria.

Resnick, S. I. (1997). Discussion of the Danish data on large fire insurance losses. *ASTIN Bulletin*, 27, 139–151.

Appendix A: Proof of Theorem 5

The residual CV in (3) can be expressed in terms of the moments of the truncated distribution. Let X be a continuous non-negative r.v. with distribution function $F(x)$, let $X1_{(X>t)}$ denote the r.v. which is equal to X if $X > t$ and equal to zero otherwise. Let $\mu_0(t) = \Pr\{X > t\}$ and $\mu_k(t) = E[X^k 1_{(X>t)}]$, $k > 0$. Throughout this paper we assume that $\mu_0(t) > 0$ for all t. Note that

$$\mu_k(t) = \mu_0(t)E(X^k \mid X > t), \tag{14}$$

hence, in particular, the residual mean and the residual variance are

$$M(t) = \mu_1(t)/\mu_0(t) - t,\; V(t) = \mu_2(t)/\mu_0(t) - (\mu_1(t)/\mu_0(t))^2,$$

and the residual CV

$$\mathrm{CV}(t) = \sqrt{\mu_2(t)\mu_0(t) - \mu_1(t)^2}/(\mu_1(t) - t\mu_0(t)).$$

Let $\{X_j\}$ be a sample of independent and identically distributed (iid) r.v.s of size n. Let $n(t) = \sum_{j=1}^{n} 1_{(X_j>t)}$ be the number of exceedances over a threshold, t. By the law of large numbers, $n(t)/n$ converges to $\mu_0(t)$. The *empirical* CV *of the conditional exceedances* is given by

$$cv(t) = cv_n(t) = \frac{n(t)}{\sum_{j=1}^{n}(X_j - t)1_{(X_j>t)}} \times \left[\frac{\sum_{j=1}^{n} X_j^2 1_{(X_j>t)}}{n(t)} - \left(\frac{\sum_{j=1}^{n} X_j 1_{(X_j>t)}}{n(t)}\right)^2\right]^{1/2}, \tag{15}$$

see (6) for a simpler expression when the r.v. are observed.

Then $cv_n(t)$ is a consistent estimator of $\mathrm{CV}(t)$ by the law of large numbers, assuming F has a finite second moment.

From Theorem 1 in Castillo et al. (2014),

$$\sqrt{n}(cv_n(t) - c_\xi) = a'(t)W(t) + O_p(1/\sqrt{n})$$

where

$$\mathrm{cov}(W(s),W(t)) \equiv M(s,t) = (\mu_{i+j}(t) - \mu_i(s)\mu_j(t))_{i,j=0,1,2},$$

and $\mu_k(t)$ are the moments of the truncated distribution (14).

$$a'(t) = (\mu_0(\mu_1 - t\mu_0), 2\mu_0(t\mu_1 - \mu_2), (-2t\mu_1^2 + t\mu_0\mu_2 + \mu_1\mu_2)) \\ /(2(\mu_1 - t\mu_0)^2\sqrt{\mu_2\mu_0 - \mu_1^2}),$$

where for simplicity dependence on t is dropped for $\mu_k = \mu_k(t)$ in the last expression. Then, the covariance function is

$$\rho_\xi(s,t) = a(s)'M(s,t)\, a(t),$$

using the conditional moments of GPD and some algebra, the result of the theorem holds.

Appendix B: R code for T_m

The following R code for T_m is used in the algorithms, see R Development Core Team (2010). See Gilleland et al. (2013) for a review of the currently available software on the generalized Pareto distribution and estimation of the extremal index.

```
#Statistic Tm of a sample given the number of thresholds m.
Tm<-function(m,sample){sam<-sample-min(sample);
     n<-length(sam);ns<-8;
     p<-round(exp(log(ns/n)/m),digits=2);
     Ws<-Ps<-Qs<-Cs<-numeric(m+1);
     for(k in 1:(m+1)){Ws[k]<-p^(k-1)};
     Ps<-1-Ws;Qs<-as.vector(quantile(sam,Ps));
     for(k in 1:(m+1))
     {Cs[k]<-sd(sam[sam>=Qs[k]]-Qs[k])/mean(sam[sam>=Qs[k]]-Qs[k])};
     cx<-(1-p)*sum(Ws*Cs)/(1-p^(m+1));xi<-(cx^2-1)/(2*cx^2);
     tm<-n*sum(Ws*(Cs-cx)^2);list(CV=cx,Tm=tm,Xi=xi)}
```