QUANTILE ESTIMATION OF HEAVY TAIL DISTRIBUTIONS AND MODEL RISKS OF EVT TECHNIQUES

Xavier CONORT

ABSTRACT

Extreme Value Theory (EVT) techniques have proven to be useful where estimation of tail quantities (extreme quantiles, small exceedance probabilities and mean excess function) is required. Such techniques have been developed in insurance and have also been implemented in finance, telecommunication, geology, and many other areas.

The aim of this paper is to give potential users a view of the model risks to which they are exposed whilst using EVT techniques.

We address misspecification risk by discussing some of the problems that may arise and suggesting useful references to test the heavy-tailedness hypothesis and to deal with challenges in data. We also suggest some alternatives when the approximations, made by statistical inference approaches in EVT, are not acceptable in practice.

In the case of system and estimation risk, we compare standard and advanced methods based on the second order framework and show that the risk can be quantified and managed for distributions with relatively good behaviour. The estimation of second order parameters provides useful information about the bias of the standard EVT techniques and can help to select the most adequate EVT estimator and threshold for the data. It also allows us to develop reduced bias techniques which significantly reduce estimation and system risk.

Such advanced techniques present their own model risks for which we will suggest areas of improvement.

INTRODUCTION

In insurance, it is essential to find a good statistical model for the largest observed historical losses. This is of particular relevance if we are required to price or build reserves for products which offer protection against catastrophic losses, such as a high-excess layer in a reinsurance treaty. Another typical problem is the determination of economic capital to constitute a cushion against irregular losses from business lines with large losses.

In such situations, Extreme Value Theory (EVT in the following) has been
recognized as being very useful, since it attempts to provide us with the best possible estimate of the tail area of the distribution, while a model chosen for its overall fit to all historical losses may not provide a particularly good fit to the large losses.

There are different ways to define the behaviour of extreme realizations and respective approaches to statistical inference:

- The Block Maxima method is based on the Fisher-Tippett theorem, which classifies the limiting distributions of Maxima. It requires us to group data into blocks, then select only the largest observation from each block. This technique is not often used in insurance since it entails discarding most of the data, and leaves some ambiguity in the choice of the block size.
- One equivalent key result is the Pickands-Balkema-De Haan Theorem. The theorem gives theoretical grounds to expect that if we choose a high enough threshold, the data above it will show Generalized Pareto behaviour.
- A third common way to estimate the behaviour of the right tail is the Hill estimator. The technique is based on the Gnedenko Theorem which states that if the distribution has a heavy tail, the tail decays like a Pareto model near infinity. The estimator is very popular since it is the EVT estimator with the lowest variance and the easiest to implement. But the estimator is only appropriate in the case of heavy tails and it exhibits relevant bias whilst having only Pareto-like tails and not a strict Pareto model.

Although those techniques have been widely used, critical voices have been raised against their imprudent use. They may indeed present strong model risks which can be categorized in 3 types:

1. They are not appropriate or efficient if application conditions are not met (mis specification risk). Data may need to be adjusted or the techniques to be adapted adequately.
2. The implementation choices (estimator, threshold, block size) are often problematic which leads to uncertainty when EVT is implemented (system risk).
3. Estimation risk can be particularly strong when the rate of convergence to EVT limiting distributions is slow.
1. Misspecification Risk

Misspecification risk arises in EVT when attempts are made to apply EVT based methods without understanding the characteristics of the loss data or the limitations of the models. First, the condition of the limit theorems may not hold. Second, there is the risk that the approximations, on which statistical inference techniques are based, are valid so far out in the tail that in practice, the amount of data needed in order to come up with reasonable results may be prohibitively large.

1.1 Theoretical Background

1.1.1 Fisher-Tippett Theorem: Limiting Distributions for Extrema

The Block Maxima method is based on the Fisher-Tippett theorem (1928) which is the fundamental result in EVT. The theorem describes the limiting behaviour of appropriately normalized sample maxima.

We consider $X_1, \ldots, X_n$ independent and identically distributed (i.i.d.) random variables with common distribution function $F$ and denote the corresponding order statistics by $X_{\cdot n} < \cdots < X_{\cdot 1}$. The weak condition of the Fisher-Tippett theorem supposes there exists sequences of constants $(a_n > 0)$ and $(b_n)$ such that the properly centered and normalized sample maxima $X_{\cdot n}$ converge in distribution to a non-degenerate limit $H$, i.e.

$$\lim_{n \to \infty} P\left( \frac{X_{\cdot n} - a_n}{b_n} \leq x \right) = H(x).$$

Then, Fisher-Tippett states that the limiting distribution $H$ belongs to one of the following extreme value distributions

- Gumbel: $H(x) = \exp(-\exp(-x))$, $x$ real
- Weibull: $H(x) = \left\{ \begin{array}{ll} \exp\left(-x^\alpha\right), & x \leq 0, \\ 1, & x > 0, \end{array} \right.$
- Frechet: $H(x) = \left\{ \begin{array}{ll} 0, & x < 0, \\ \exp\left(-\frac{x}{\alpha}\right), & x \geq 0, \end{array} \right.$

$F$ is said to belong to the max-domain of attraction of one of the three distributions: Weibull, Gumbel or Frechet.
A unified formula for the previous three limiting distributions is provided by the Von Mises (1936) representation: the generalized extreme value (GEV) which is very useful in practice since it nets three types of limiting behaviour in one framework.

\[
H_x(x) = \begin{cases} 
\exp\left(\frac{-(1+y)x}{y}\right) & \text{if } y \neq 0 \text{ and } 1+yx > 0 \\
\exp(-\exp(-x)) & \text{if } y = 0 
\end{cases}
\]

if \( \gamma < 0 \) (Weibull): \( \Phi_{\gamma,x}(x) = H_x(-x+1)/\gamma \)
if \( \gamma = 0 \) (Gumbel): \( \Lambda(x) = H_x(x) \)
if \( \gamma > 0 \) (Frechet): \( \Phi_{\gamma,x}(x) = H_x((x-1)/\gamma) \)

The real parameter \( \gamma \), which appears in the Von Mises representation, is referred to as the Extreme Value Index (EVI) of \( F \). It rules the behaviour of the right tail of \( F \) and helps to indicate the size and frequency of extreme events. The larger the index, the heavier the tail.

In practice, most common continuous distribution functions satisfy the weak condition of the Fisher-Tippett theorem quite naturally. Four kinds of tails can be seen:

- Short-tailed distributions with finite right endpoint, such as the uniform, beta and reversed Burr distributions belong to the Weibull domain of attraction. This class generally is of lesser interest in insurance applications.

- Distributions possessing an exponential tail, having finite right endpoint or not, belong to the Gumbel domain. Distributions in this max-domain of attraction include the normal, exponential, gamma and lognormal distributions. They are often called light tailed distributions (or medium tailed). Particular mention should be made of the lognormal distribution which has a much heavier tail than the normal distribution. It has historically been a popular model for loss severity distributions in insurance; however, it is not technically a heavy tailed distribution.

- Pareto-type (or heavy tailed) distributions belong to the Frechet domain and have polynomially decaying tails. Such tails can be seen in insurance (loss data), finance (stock returns), telecommunication (file sizes, waiting times), geology (diamond values, earthquake magnitudes), and many others. The class of distribution includes the Pareto, Burr, Log-gamma, Cauchy and Student distributions as well as various mixture models.
- Super-heavy tails (log Pareto, log Weibull etc.) are distributions with so much weight in the tail that they do not belong to any max-domain of attraction (EVT=infinite). More details on such class of tails can be found in Fraga Alves, de Haan and Neves (2009).

The Block Maxima method was suggested by Gumbel (1958). Under Gumbel’s model, the sample of size \( n \) is divided into \( k \) sub-samples of size \( r \). Next, the maximum of the \( r \) observations in each of the \( k \) sub-samples is considered, and one of the three extremal models (Weibull, Gumbel and Frechet) is fitted to the sample of those \( k \) maximum values.

Nowadays, whenever using this approach, still quite popular in environmental sciences, it is more common to fit to the data an extreme value distribution function \( H(x - \lambda)/\delta \), with \( H \) given in the Von Mises representation and \( (\lambda, \delta, \gamma) \) unknown location, scale and “shape” parameters.

Computational details on the estimation of \( (\lambda, \delta, \gamma) \) and useful references can be found in Gomes, Canto e Castro, Fraga Alves and Pestana (2008).

1.1.2 Pickands-Balkema-De Haan theorem: limiting distributions for the tails

The Generalized Pareto Distribution (GPD) fitting technique is considered to be more useful for practical applications than the Block Maxima method, due to its more efficient use of data.

It is based on the Pickands-Balkema-De Haan Theorem (1974, 1975) that states that, under the same conditions as for the Fisher-Tippett theorem, the GPD is the limiting distribution for the distribution of the excess losses, as the threshold tends to the right endpoint. Then, we can find an appropriate normalizing function \( \sigma(u) \) such that

\[
\lim_{u \to x_0} \sup_{u \leq x \leq x_0} \left| F_x(x) - G_{x,\sigma(u)}(x) \right| = 0
\]

with \( G_{x,\sigma}(x) = \begin{cases} 1 - (1 + x/\sigma)^{-1/\gamma} & \text{if } \gamma \neq 0 \\ 1 - \exp(-x/\sigma) & \text{if } \gamma = 0 \end{cases} \) the GPD

\[
F_x(x) = P[X - u \leq x | X > u] = \frac{F(x + u) - F(u)}{1 - F(u)}
\]

the distribution of excess losses

for \( 0 \leq x \leq x_0 - u \) and \( x_0 \) the finite or infinite right endpoint.

In practice, we restrict our attention only to the observations that exceed a certain high threshold \( u \), and assume that the excess distribution above this threshold may be taken to be exactly GPD for some \( \gamma \) and \( \sigma \).
The fitting method commonly used is the Maximum Likelihood estimation (MLE) and the choice of the threshold is basically a compromise between:

- choosing a sufficiently high threshold so that the asymptotic theorem can be considered to be essentially exact and
- choosing a sufficiently low threshold so that we have sufficient material for the estimation of the parameters.

The modelling approach was developed in Davison and Smith (1990) and other papers by these authors. For the reader interested in a detailed description, McNeil (1997) provides also an extensive example of its application to Danish data on large fire insurance losses.

1.1.3 Gnedenko theorem: limiting distribution for heavy tails

The Hill estimator is based on a result found by Gnedenko (1943) who showed that for distributions which belong to the Frechet max-domain of attraction ($\gamma > 0$) and under the same conditions as the previous theorems, the tail of $F$ decays like a power function and equivalently the tail quantile function $U$ increases like a power function near infinity.

\[
U(x) = x^\gamma L(x) \quad \text{with} \quad \gamma > 0 \\
\text{with} \quad U(t) = F^{\gamma} - (t - 1)/t \quad \text{with} \quad t > 1 \quad \text{and} \quad F^{\gamma}(x) = \inf\{y : F(y) \geq x\}
\]

and $L$ a slowly varying function at infinity such that $L(tx) / L(t) \to 1$ when $t \to \infty$.

The Hill estimator has been proposed by Hill (1975) and built on a first order approximation of the previous result. He makes the approximation that for a Pareto-type distribution, above a certain threshold, the relative excesses behave as data from a strict Pareto distribution.

\[
U(tx)/U(t) = x^\gamma \quad \text{or equivalently} \quad F(tx)/F(t) = x^{-\gamma} = x^{-\alpha} \quad \text{with} \quad F = 1 - F
\]

Then, the fitting method is the MLE which conducts to a simple functional form for the estimator of the EVI

\[
\gamma_{\text{Hill}} = \frac{1}{k} \sum_{i=1}^{k} \left\{ \log X_{n-i+1}^* - \log X_{n-i}^* \right\}
\]

and $X_1^* \leq X_2^* \leq \ldots \leq X_n^*$ the order statistics

Its properties have been thoroughly studied by several authors among whom we can mention de Haan and Peng (1998).
1.2 Application issues

The weak condition of the Fisher-Tippett theorem, that the distribution of appropriately normalized sample maxima converge to a non-degenerate distribution is known to be satisfied by most common continuous distribution functions. Tails which do not belong to any max-domain of attraction are rare. Examples are super-heavy tails (log Pareto, log Weibull etc.) for which the Extreme Value Index is infinite.

In contrast, estimations performed under the assumption that data are independent and identically distributed (i.i.d.) can be over-optimistic. Indeed, we often deal in practice with

- losses which are not of the same general type (acquisition, change in exposure over time, contaminated data)
- loss amounts which are not comparable (inflation)
- losses dependent on one another (financial return series)
- and a future which holds some unexpected catastrophic losses that are not reflected by the historical loss data (switching regime, changes in regulation and legislation)

An additional assumption for the Hill estimator technique is that the model underlying the data belongs to the Frechet max-domain. The GPD approximation technique does not require such an assumption since the limiting distribution nests all limiting behaviours in a unified framework. However, statistical inference is improved if we make a priori choice.

One less known concern in EVT is the rate of convergence in its limit theorems. There exists no equivalent of the Berry-Esseen theorem that, under broad conditions, gives a rate of convergence of the order of $1/\sqrt{n}$ in the Central Limit Theorem (CLT).

In EVT, the rate of convergence depends strongly on the right tail of $F$. If $F$ has only (G)PD-like tails, the approximations, made in practice that the excess distribution above a finite threshold $u$ may be taken to be exactly GPD or the relative excess distribution to a Pareto, can be unacceptable for finite samples. The resulting bias can constitute a serious problem and produces large errors in the estimation of extreme quantiles and other tail quantities based on the Extreme Value Index.

Examples of distributions with such bad behaviour are given by Degen, Embrechts and Lambrigger (2006). They show that for the G-h distribution, convergence is extremely slow and that for the very popular Log-normal and Log-gamma distributions, convergence
is very slow. In contrast, for distributions like the Normal, Student and Weibull, convergence is at a reasonably fast rate.

In the case of the Pareto approximation, this mostly happens when the slowly varying part in the Gnedenko theorem disappears at a very slow rate. Log-gamma and G-h distributions have a slow decay term (such as a log x) which prevents the asymptotic convergence towards the Pareto distribution. Examples of fat tailed distributions with relatively good behaviour are given by the Hall-Welsh class of models which includes the Generalized Pareto, Burr, Frechet, Cauchy, and Student distributions. Their tails have a second order expansion where the first and second terms have the same functional form and are characterized by the existence of constants $C$, $\gamma$, $\rho$ and $\beta$ such that:

\[
F_{\text{Hall-Welsh}}(x) = \left( \frac{x}{c} \right)^{\gamma/h} \left( 1 + \frac{\beta}{\rho} \left( \frac{x}{c} \right)^{\gamma/h} + o\left( x^{\gamma/h} \right) \right) \quad \text{as} \quad x \to \infty \quad \text{with} \quad C > 0, \beta \neq 0 \text{ and } \rho < 0
\]

where $\gamma$ is the first order parameter (or EVI), $\rho$ a shape second order parameter (also called the rate of convergence, as we will see later) and $\beta$ a scale second order parameter.

One way to illustrate those behaviour is to use the Pareto quantile plot which is a goodness-of-fit tool to assess the Pareto approximation. Instead of using directly the classical QQ plot, the plot is an exponential quantile plot based on the log-transformed data, since log-transformed Pareto random variables are exponentially distributed.

- If the data originate from a strict Pareto distribution, the Pareto quantile plot will show a straight line pattern, the slope of which is given by the Extreme Value Index.
- If the data originate from a Pareto-type distribution, the Pareto quantile plot will be linear but only in some of the largest observations.
Graph 1 – Theoretical Pareto quantile plot with true quantiles of Student, Burr and Log-gamma distributions

If the data are well modelled by a Log-gamma distribution, the Pareto quantile plot will be linear only in extremely high observations, which means that in practice, the Pareto approximation may be unacceptable for such data.

1.3 Useful references to deal with misspecification risk

We suggest here useful references to deal with some of the problems discussed earlier.

1.3.1 Risk that the type of tail is not well detected

The choice of the most appropriate tail type for the underlying distribution function has become a usual practice since statistical inference is improved if we make a priori choice. Indeed, the detection of the intermediate case of distributions in the Gumbel domain (light tails) is of particular interest because of the simplicity of inference within this domain. It also allows us to identify if the use of the Hill estimator is appropriate.

Graphical procedures

The QQ-plot against the exponential distribution is one of the most popular procedures to detect heavy-tailedness. It examines visually the hypothesis that the losses come from an exponential distribution, i.e. from a distribution with a light tail. The quantiles of the empirical distribution function on the x-axis are plotted against the quantiles of the exponential distribution function on the y-axis.
- The points should lie approximately along a straight line if the data are an i.i.d. sample from an exponential distribution (Gumbel domain).
- A concave departure from the ideal shape indicates a heavier tailed distribution (Frechet domain)
- whereas convexity indicates a shorter tailed distribution (Weibull domain).

This technique cannot be trusted fully as pointed out by McNeil (1997). Even data generated from an exponential distribution may sometimes show departures from typical exponential behaviour.

A further useful graphical tool is the plot of the sample mean-excess function. It is an empirical estimate of the mean excess function which describes the expected overshoot of a threshold given that exceedance occurs.

- If the points show an upward trend, then this is a sign of heavy tailed behaviour (Frechet domain).
- Exponentially distributed data would give an approximately horizontal line (Gumbel domain)
- Data from a short tailed distribution would show a downward trend (Weibull domain).

As for the QQ-plot, there are some problems associated with its use. Ghosh and Resnick (2009) pointed out that the plot is not a very effective diagnostic tool for discerning the model in the case of a lognormal distribution and is not appropriate for an Extreme Value Index higher than 1 (mean value infinite).

**Graph 2 – Plots of the sample mean-excess function for two data samples of size n=5000 from the same Lognormal distribution (meanlog=0, sdlog = 1)**
Testing procedures statistics
Testing procedures statistics are useful complements to the graphical procedures.

Gumbel vs Frechet or Weibull domains
Alves and Neves (2006) gave an overview of test statistics for Gumbel domain against Frechet or Weibull domains. They recommend some recent semi-parametric approaches based on location/scale invariant statistics built on the observations from the sample lying above a random threshold (PORT approaches). Three procedures are suggested by them: the NPFA-test statistic (ratio between the Maximum and the Mean of Excesses); the Greenwood-type statistic (Gt-test) and the Hasofer and Wang statistic (HW-test). According to an extensive simulation study, they conclude that:

- The Gt-test is shown to good advantage when testing the presence of heavy-tailed distributions in demand
- The HW-test is the most powerful test under study concerning alternatives in the Weibull domain of attraction while the Gt-test barely detects small negative values of EVI
- The NPFA-test tends to be a conservative test

Frechet vs Gumbel domain
Goegebeur, Beirlant and de Wet (2006) introduced kernel goodness-of-fit statistics for assessing whether a sample is consistent with the Pareto-type model (Frechet domain).

They exploit the close link between the Pareto-type and the exponential model, put some of the available goodness-of-fit procedures for the latter in a broader perspective and suggest the use of Modified Jackson and Lewis test statistics.

Graph 3 – Plot of Modified Jackson and Lewis test statistics for a data sample of size $n=5000$ from a Burr distribution (shape1=2, shape=1, scale=10)
In addition to providing a tool to assess the hypothesis that the underlying distribution is of Pareto-type, we can notice that the test statistics also enable us to select a range of appropriate thresholds for the Extreme Value Index estimation.

**Super-heavy tails vs Frechet domain**

Alves, de Haan and Neves (2008) suggested test statistics to assess models with potential super-heavy tails (log Pareto, log Weibull etc.) for which no EVT technique works.

**1.3.2 Risk that data are dependent on one another**

In the case of dependent data, the (G)PD-based tail estimator is still asymptotically valid, but provides much less stable results compared to the i.i.d. case.

McNeil and Frey (2000) suggested, for heteroscedastic financial return series, the use of conditional EVT. They propose a two steps procedure: first filtering the returns through a more or less complex GARCH model, and second, estimating the tail parameters using the assumption of i.i.d. data.

**1.3.3 Risk that data are censored**

Data sets with censored extreme data may typically occur in insurance when reported payments cannot be larger than the maximum payment value of the contract. Some adaptations of the Hill estimator are proposed by Beirlant, Delafosse and Guillou (2007) and used to build extreme quantile estimators in the case of censoring.

**1.3.4 Risk that data are contaminated**

Both Hill and GPD techniques are constructed on Maximum Likelihood estimators based on specific parametric models (Pareto and Generalized Pareto distributions) which are fitted to excesses over large thresholds. Maximum likelihood estimators are sensitive to a few particular observations. In order to overcome the problem of a few abnormal observations, Vandewalle, Beirlant, Christmann and Hubert (2007) proposed a more robust estimator of the tail index which combines a refinement of the Pareto approximation with an integrated squared error approach on partial density component estimation.

**1.3.5 Risk that data are shifted**

Hill technique can be very sensitive to changes in location since a shift will affect the value of the relative excesses values and then the rate of convergence of their distribution to a Pareto.
In the literature, it has been sometimes suggested to subtract a random quantity, usually the minimum of the sample. This adjusted data set has the advantage of working with more non-negative values. However, Gomes, Alves and Santos (2007) pointed out that this adjustment, for distributions with infinite left and right end points (such as the Student distribution), can lead to large flat zones in the Hill estimates path, which strongly underestimate the Extreme Value Index (EVI) (see the Hill(0) path in the following graph) and can lead in practice to misleading results.

They advise:
- for distributions with infinite left and right end points, to adjust the data with a shift induced by the median or the first empirical quartile or decile
- for distribution with finite left end points, to subtract the minimum of the sample

Graph 4 – Hill estimates paths for data samples of size n=5000 from a centred Student distribution with 2 degrees of freedom and shifts equal to empirical quantiles (p=0, 0.1, 0.25, 0.5) in function of k (nb of top order statistics).

1.3.6 Risk that the (G)PD approximation is unacceptable

We saw earlier that Degen, Embrechts and Lambrigger (2006) pointed out that for some distributions, convergence of excess (relative excesses) distribution to the GPD (Pareto distribution) may be very slow. If the GPD or Pareto approximations are found to be unacceptable in practice, one should consider alternatives.
Parametric distribution fitting

Parametric distribution fitting is one possible alternative. In this case, a quantile-based method for the estimation of the parameters is preferred to commonly used estimation methods such as the method of moments and maximum likelihood estimation (MLE) since those methods are potentially less accurate for fitting the tails of a distribution.

Such an approach is recommended by Dutta and Perry (2006) to estimate the operational risk of financial institutions. They conducted an experiment using financial institutions internal loss data collected under the 2004 Loss Data Collection Exercise (joint effort of US banking regulatory agencies) to comparatively analyse various approaches and used data from seven institutions. To model the severity distribution, they used three different techniques: parametric distribution fitting, the GPD technique and capital estimation based on non-parametric empirical sampling. They considered one- and two-parameter distributions to model the loss severity: exponential, gamma, generalized Pareto, loglogistic, truncated lognormal and Weibull distributions. They also used four-parameter distributions such as the Generalized Beta Distribution of Second Kind (GB2) and the G-and-h distribution.

In their experiment, G-and-h distribution was found to perform the best and yielded the most realistic and least varying capital estimates across institutions at the enterprise, business line, and event type levels. In contrast, the EVT approaches gave the most unreasonable capital estimates with the most variation of all of the methods.

The findings from Dutta and Perry experiment are explained by Degen, Embrechts and Lambrigger (2006). They show that for the g-and-h distribution, convergence of the excess distribution to the generalized Pareto distribution (GPD) is extremely slow. Therefore quantile estimation using EVT leads in practice to inaccurate results if data are well modelled by a g-and-h distribution.

Penultimate estimation

We saw that for distributions such as the Log-gamma, estimating the EVI at infinity can be pointless if the aim is to estimate a finite quantile (or VaR). Degen and Embrechts (2009) suggested considering instead local heavy-tailedness (or penultimate estimation). Locally, i.e. at any point t0, the tail quantile function U may be interpreted as an exact Pareto model \( U(t) = ct^{\gamma} \) with tail-index \( \gamma = \phi'(\log t_0) \) and some (unknown) \( c = c(t_0) \), by setting \( U(t) = \exp(\phi(\log t)) \).
Then, the VaR can be estimated in a similar way as the Weissman technique but where the EVI parameter at infinity is replaced by the estimate of the local slope of $\phi$ using local (quadratic) regression (with a tricube weight function and smoothing parameter of $3/4$).

2. SYSTEM RISK

Once the type of tail has been detected and the data or the estimator adjusted, companies are exposed to a second risk, the system risk.

Indeed, EVT implementation may be uncertain since, as we saw earlier,

- there are different approaches to statistical inference on extreme values and
- an additional practical difficulty involves choosing an appropriate threshold.

The impact of uncontrolled decisions by the actuary can be in fact significant over time: using different methods or thresholds can result in materially different estimates of a company exposure.

In order to mitigate the system risk, one can choose to impose a standardized approach to apply in all cases and then, restrict the choice of the estimator and the choice of the threshold above which GPD and Pareto distributions are fitted (GPD and Hill methods).

This approach would ensure consistency and is easily implemented. However, in EVT, a push-of-the button attitude may lead to non optimal choices and increases significantly estimation risk.

- Among classical EVT techniques, none performs uniformly best and can be recommended in all cases.
- Arbitrary threshold setting may lead to strong overestimation (in particular, in the case of the Hill estimator). Indeed, the choice of a threshold should be the result of a bias-variance trade-off which strongly depends on the model underlying the data sample and should not be arbitrary.

An alternative to the push-of-the button is to set up an internal control system which would provide the actuary with a framework for the choice of the techniques and thresholds. Graphical tools to justify those choices can be part of this framework and easily implemented. However, one should not trust them fully and companies are also exposed to differences in their interpretation. In addition, this approach cannot be easily audited.

We suggest in this paper the implementation of more advanced techniques which
may help to choose the EVT estimator and the threshold more systematically. Such techniques consist in estimating the bias/variance of each estimator at different thresholds and selecting the estimator-threshold couple which optimizes the estimation error.

They allow a more objective choice which may be easily tracked. The outputs can also be easily shared among users of a same group and enhance the overall experience.

Those advanced techniques are however difficult to implement and some are relatively recent and not sufficiently tested in practice.

2.1 Relative performance of the standard estimators

We stated earlier that among classical EVT techniques, none performs uniformly best. We illustrate it here through Monte Carlo simulations. We will show later (3.1) that this statement is also supported by the theory.

We implemented Monte Carlo simulations of 200 finite samples of size \( n=5000 \) from Frechet, Burr, GP and Student distributions with different combinations of the parameters \( \gamma, \rho \) and \( \beta \) which characterize the distributions in Hall Welsh’s class such that:

\[
F_{\text{Hall-Welsh}}(x) = \left( \frac{x}{c} \right)^{\gamma \rho} \left( 1 + \frac{\beta}{\rho} \left( \frac{x}{c} \right)^{\rho + 1} \right) \quad \text{as } x \to \infty \quad \text{with } C > 0, \beta \neq 0 \text{ and } \rho < 0
\]

The parameters of the Frechet, Burr and Student distributions were set following the Table 1 provided by Caeiro and Gomes (2008):
### Table 1 – First and second order parameters of some distributions in Hall-Welsh’s class

We computed the estimation of the Extreme Value Index using the standard techniques and compared their relative standard errors assuming that the optimal threshold is known (threshold with the lowest standard error).

---

**Burr parent (EVI = 1/shape1/shape2 & Rho = -1/shape1)**

<table>
<thead>
<tr>
<th>EVI values</th>
<th>0.25</th>
<th>0.5</th>
<th>0.75</th>
<th>1</th>
<th>1.25</th>
<th>1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rho values</td>
<td>-0.25</td>
<td>GPD</td>
<td>Moment</td>
<td>Moment</td>
<td>Moment</td>
<td>Moment</td>
</tr>
<tr>
<td></td>
<td>-0.5</td>
<td>HILL</td>
<td>GPD</td>
<td>GPD</td>
<td>GPD</td>
<td>GPD</td>
</tr>
<tr>
<td></td>
<td>-0.75</td>
<td>HILL</td>
<td>Moment</td>
<td>Moment</td>
<td>Moment</td>
<td>Moment</td>
</tr>
<tr>
<td></td>
<td>-1</td>
<td>HILL</td>
<td>Moment</td>
<td>Moment</td>
<td>Moment</td>
<td>Moment</td>
</tr>
<tr>
<td></td>
<td>-1.25</td>
<td>HILL</td>
<td>HILL</td>
<td>HILL</td>
<td>Moment</td>
<td>GPD</td>
</tr>
<tr>
<td></td>
<td>-1.5</td>
<td>HILL</td>
<td>HILL</td>
<td>HILL</td>
<td>HILL</td>
<td>GPD</td>
</tr>
<tr>
<td></td>
<td>-1.75</td>
<td>HILL</td>
<td>HILL</td>
<td>HILL</td>
<td>HILL</td>
<td>GPD</td>
</tr>
<tr>
<td></td>
<td>-2</td>
<td>HILL</td>
<td>HILL</td>
<td>HILL</td>
<td>HILL</td>
<td>GPD</td>
</tr>
</tbody>
</table>

**Student parent (EVI = 1/ν & Rho = -2/ν)**

<table>
<thead>
<tr>
<th>Degrees of freedom (ν)</th>
<th>4</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>EVI values</td>
<td>0.25</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>Rho values</td>
<td>-0.5</td>
<td>-1</td>
<td>-2</td>
</tr>
<tr>
<td>HILL</td>
<td>Moment</td>
<td>GPD</td>
<td></td>
</tr>
</tbody>
</table>

---

**Frechet parent (EVI = 1/shape & Rho = -1)**

<table>
<thead>
<tr>
<th>EVI values</th>
<th>0.5</th>
<th>1</th>
<th>1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rho values</td>
<td>GPD</td>
<td>GPD</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RSE &gt; 25%</th>
<th>10% &lt; RSE &lt; 25%</th>
<th>5% &lt; RSE &lt; 10%</th>
<th>RSE &lt; 5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moment</td>
<td>GPD</td>
<td>GPD</td>
<td></td>
</tr>
</tbody>
</table>

---

**Table 2 – Standard EVT technique for the estimation of EVI with the lowest RSE according to Monte Carlo simulations of samples of size n=5000 from models in Hall-Welsh’s class**

We can notice in table 2 that none of the standard estimators performs uniformly best:
The GPD estimator outperforms when $\gamma$ and $\rho$ are close to $\gamma + \rho = 0$
- Hill performs best while $\gamma < \rho$.
- The Moment estimator is a good estimator for $\gamma$ values between 0.5 and 0.75. For small $|\rho|$ values, our simulation shows also a superiority of the Moment estimator over the GPD and Hill techniques.

We also performed Monte Carlo simulations for the estimation of the VaR at a 99.9% confidence level and compared the relative performances of the standard estimators, still assuming that the optimal threshold is known.

<table>
<thead>
<tr>
<th>Rho values</th>
<th>EVI values</th>
<th>Bur parent</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.25</td>
<td>GPD</td>
<td>0.25</td>
</tr>
<tr>
<td>-0.5</td>
<td>GPD</td>
<td>0.5</td>
</tr>
<tr>
<td>-0.75</td>
<td>Moment</td>
<td>0.75</td>
</tr>
<tr>
<td>-1</td>
<td>GPD</td>
<td>1</td>
</tr>
<tr>
<td>-1.25</td>
<td>HILL</td>
<td>1.25</td>
</tr>
<tr>
<td>-1.5</td>
<td>HILL</td>
<td>1.5</td>
</tr>
<tr>
<td>-1.75</td>
<td>HILL</td>
<td>GPD</td>
</tr>
<tr>
<td>-2</td>
<td>HILL</td>
<td>GPD</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Degrees of freedom ($\nu$)</th>
<th>EVI values</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.25</td>
</tr>
<tr>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>-0.5</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>-2</td>
<td>-2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fréchet parent</th>
</tr>
</thead>
<tbody>
<tr>
<td>EVI values</td>
</tr>
<tr>
<td>0.5</td>
</tr>
<tr>
<td>GPD</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>GPD</td>
</tr>
<tr>
<td>1.5</td>
</tr>
<tr>
<td>GPD</td>
</tr>
</tbody>
</table>

Table 3 – Standard EVT technique for the estimation of VaR(99.9%) with the lowest RSE according to Monte Carlo simulations of samples of size $n=5000$ from models in Hall-Welsh’s class

As for the EVI estimation, none of the standard estimators performs uniformly best, except that the region of over performance for the GPD technique is slightly larger.

2.2 Selection of the thresholds

The issue of how to select a good, or, if possible, an optimal threshold is also problematic. It can present a system risk since the techniques used in practice are mostly visual and exposed to subjective interpretation.

The mean-excess plot that we saw earlier to detect heavy tails, is also used to detect a good threshold in the GPD case by looking for a reasonably straight line in the plot.

It is also common to check whether the (GPD or PD) estimates of the EVI parameter are stable in function of the number of top order statistics (k) used for the estimation.
Graph 5 – Hill estimates path for a data sample of size $n=5000$ from a Burr distribution ($\text{shape}1=2$, $\text{shape}=1$, $\text{scale}=10$) and $k$ (the number of top order statistics) using the test statistics of Graph 3

In Graph 5, we can notice that if the rate of convergence in the Gnedenko theorem is slow, this threshold stability may not be visible in the Hill estimates path for a given data set. The same issue may arise when applying the Pickands-Balkema-de Haan theorem and then fitting a GPD to the excesses over a high, positive threshold.

To validate the model once an a priori choice of the threshold $u$ has been made, the Pareto quantile plot or the plot of empirical excesses distribution vs the fitted GPD are useful goodness of fit graphs. The kernel goodness-of-fit statistics introduced by Goegebeur, Beirlant and de Wet (2006) and that we presented earlier, is also an interesting procedure to validate the choice of the threshold in case of the Hill estimator.

Graph 6 – Pareto quantile plot for a data sample of size $n=5000$ from a Burr distribution ($\text{shape}1=2$, $\text{shape}=1$, $\text{scale}=10$) and the threshold $u$ selected such that the number of top order statistics $k$ is equal to 100
In order to adopt a more systematic and optimal way, several authors have suggested estimating and minimizing the standard error. Danielsson, de Haan, Peng and de Vries (2001) suggested a bootstrap approach. We will see that other techniques using the second order framework can also be implemented.

3. ESTIMATION RISK

We discuss, in this section, how the estimation risk can be approximated using a second order framework. This information can be used to minimize the standard error in order to select the optimal couple of estimator and threshold. We will see that an even more powerful use of the second order framework is the development of reduced bias estimators which not only mitigate the system risk by making the choice of the threshold easier but also reduce the estimation risk.

3.1 Distributional behaviour of the standard estimators under the second order framework

We saw, that according to Gnedenko (1943) for distributions which belong to the Frechet max-domain of attraction ($\gamma > 0$), the tail quantile function $U$ increases like a power function near infinity.

$$U(x) = x^\gamma L(x) \text{ with } \gamma > 0 \text{ and } L \text{ a slowly varying function at infinity}$$

We are interested here in the rate at which the Hill’s approximation $U(tx)/U(t) - x^\gamma$ tends to 0 as $t$ tends to infinity or equivalently how fast $L(tx)/L(t)$ tends to 1.

For determining such a rate of convergence, we need to assume the second order condition introduced by Geluk and de Haan (1987), which holds for a large class of heavy tail models. In this condition we assume that there exists a function $A$ such that $|A(t)|$ is of regular variation with index $\rho$ such that:

$$\lim_{t \to \infty} \frac{\ln U(tx) - \ln U(t) - x \ln x}{A(t)} = \begin{cases} x^\rho - 1 & \rho < 0 \\ \rho \ln x & \rho = 0 \end{cases}$$

For distributions in Hall-Welsh’s class, $\rho$ is strictly negative and the second order condition holds when $A(t) = \gamma t^\rho$. For distributions with bad second order behaviour such as the Log-gamma and the G-h distributions, $\rho$ is equal to 0.
According to de Haan & Peng (1998), Gomes and Rodrigues (2008) and other papers by the same authors, under the validity of the second order condition, and for intermediate levels \(k\) (the number of top order statistics) such that \(\sqrt{k} A(n/k) \to \lambda\), finite, as \(n \to \infty\), we may write for the Hill, Moment and GPD estimators of the EVI that:

\[
\hat{\lambda}_{HILL,k} = \gamma + \frac{1}{\sqrt{k}} Z_k + \frac{A(\frac{q}{k})}{1-\rho} + o_p(A(\frac{q}{k}))
\]

\[
\hat{\lambda}_{MOMENT,k} = \gamma + \frac{\sqrt{1+\gamma^2}}{\sqrt{k}} L_k + \frac{(\gamma(1-\rho)+\rho)A(\frac{q}{k})}{\gamma(1-\rho)^2} + o_p(A(\frac{q}{k}))
\]

\[
\hat{\lambda}_{GPD,k} = \gamma + \frac{1+\gamma}{\sqrt{k}} M_k + \frac{(1+\gamma)(\gamma+\rho)A(\frac{q}{k})}{\gamma(1-\rho)(1-\rho+\gamma)} + o_p(A(\frac{q}{k}))
\]

with \(Z_k, M_k\) & \(L_k\) asymptotically standard normal.

In the case of Hall-Welch’s models, we know that \(A\) has the functional form \(A(t) = \gamma \beta t^\rho\), which allows us to compare easily the relative performances (variance and bias) of the estimators in function of the parameters \(\gamma, \rho\) and \(\beta\) and the number of the top order statistics \(k\).

For a Burr distribution (for which \(\beta=1\)), we observe the following:

- For low values of EVI (or \(\gamma\)), the GPD estimation variance is large and prevents the method from performing well, while for high values of \(\gamma\), the variance is close to the variance of Hill’s EVI estimates.
Graph 8 – Theoretical Hill, Moment & GPD 2nd order bias of EVI estimations for Burr models with different values of EVI in function of the value taken by ρ

- The GPD estimation is unbiased when γ+ρ=0 and lower than Hill’s bias around these values. For small values of γ, the GPD estimation can have a strong negative bias.

- EVI estimates of the Moment estimator have a lower variance than the GPD method and its bias is particularly low for γ values between 0.5 and 0.75. But it has a strong negative bias for smaller values of γ. Therefore it may be a good candidate for the EVI estimation but only for certain cases.

For other Hall-Welsh models such as Frechet or Student distributions, we find similar results. Only the scale, which depends on β, is different.

These theoretical results based on the 2nd order approximation support our previous findings obtained through Monte Carlo simulations in Hall-Welsh’s class that among classical EVT techniques, none performs uniformly best and can be recommended in all cases.

Combined with a prior estimation of the second order parameters, it can be used to estimate the errors of the standard estimators and then select the estimator-threshold couple which optimizes the estimation error.

However, it is important to keep in mind our assumption that the model underlying the data belongs to a Hall-Welsh class which may not be satisfied and even if the optimal level of standard error is reached, its level can still be unacceptable.

Particular mention should also be made that:

- the previous theoretical distributional behaviours are based on a second
order framework and the 2\textsuperscript{nd} order approximation proposed holds only for small values of \(k\).

when the rate of convergence is slow, the true standard error function has a “very peaked” shape (very short stability regions around the target value) which can make the choice of the optimal threshold difficult in practice.

Graph 9 – Hill, Moment & GPD’s Relative Standard Errors of EVI estimations in function of \(k/n\) for two Burr models of size \(n\)=5000 (Set1: EVI=1.5 and Rho=-1.5 / Set2: EVI=0.75 and Rho=-0.5)

**Continuous lines:** Theoretical pattern of RSE based on the 2nd order approximation

**Dashed lines:** Simulated RSE via a Monte Carlo simulation of 200 samples

3.2 Estimation of the second order parameters

3.2.1 Estimation of the rate of convergence \(\rho\)

We present here a technique to estimate the rate of convergence \(\rho\) which was introduced by Alves, Gomes and de Haan (2003) and has been recommended by several authors.

The estimator of \(\rho\) is parameterized by a tuning (or control) parameter \(\tau\).

\[
\hat{\rho}_\tau(k) = -\min(0; 3(\hat{T}_n^{(\tau)}(k) - 1)/(\hat{T}_n^{(\tau)}(k) - 3))
\]

with
If $\tau$ is chosen adequately, the estimator of $\rho$ is unbiased, asymptotically normal, and shows highly stable sample paths as functions of $k$, for a wide range of large $k$-values whenever the second order condition holds.

In practice, similar algorithms have been advised to select the tuning parameter, where the choice of the tuning parameter is limited to positive values $\tau = 0$ and $\tau = 1$ and the estimation of $\rho$ is computed at the value $k_1 = n^{0.995}$, not chosen in any optimal way.

Algorithm 1 (Gomes, Rodrigues, Pereira and Pestana (2008))

1. Given a sample of $n$ positive observations $(X_1, X_2, \ldots, X_n)$, plot for $\tau = 0$ and $\tau = 1$ the estimates $\hat{\rho}_1(k)$
2. Consider $\{\hat{\rho}_1(k)\}_{k \in K}$ for integer values $k \in [0^{0.999} n^{0.999}]$ and compute their median, denoted, $\rho^*$. Next choose the tuning parameter $\tau^* = \arg\min_{\tau} \sum_{k \in K} (\hat{\rho}_1(k) - \rho^*)^2$.
3. Work then with $\hat{\rho}_1 = \hat{\rho}_1(k_1)$ and $\hat{\beta}_1 = \hat{\beta}_1(k_1)$, where $k_1 = n^{0.995}$

Several authors noticed some discrepancies in this algorithm in particular when $\rho$ has small absolute values. Caeiro and Gomes (2008) pointed out that the parameter $\tau$ is not necessarily non-negative but real. Under a third order framework and under other conditions fulfilled by the classical models such as Frechet, Student, Burr and the GPD, they prove that the optimal candidate of $\tau$ is a function of $\rho$ and $\beta$ and a third order parameter $\beta'$. 
Graph 10 – Theoretical optimal value of the tuning parameter $\tau$ in function of the rate of convergence $p$ based on a third order approximation

Such an optimal value is difficult to compute in practice, but the theoretical result should encourage us to enlarge the range of possible candidates to negative values.

On the basis of this result, Caeiro and Gomes advised drawing a few sample paths with different values of $\tau$, as functions of $k$, selecting the $\tau$ value which provides the highest stability for large $k$.

Graph 11 – Rho estimates paths with different values of the tuning parameter $\tau$ for a data sample of size $n=5000$ from a Burr distribution ($\text{shape}_1=2$, $\text{shape}=1$, $\text{scale}=10$)
In line with this recommendation, we tested another algorithm (Algorithm 2) where we
1. considered 36 potential values from -2 to 1.5 as candidates for $\tau$ and
2. based our stability criterion for a region of large $k$ values varying according
to a prior estimation of the estimator variance.

![Graph 12 – Theoretical Rho estimator Variance for Hall-Welsh models with
different values of $\rho$ ($\beta=1$)](image)

The second improvement suggested is driven by the observation that for small
absolute values of $\rho$ ($\rho>0.5$), the region of $k$ values for which the estimator has a high
stability is significantly larger than the region suggested in the algorithm 1. By enlarging
the region, we aim to use more statistical material in order to improve the quality of the $\rho$
estimation.

On the basis of this observation, we suggest evaluating an enlarged region $K^{**}$
using a prior estimation of $\rho$ & $\beta$ that we obtained with the region $K^*$ suggested in
algorithm 1. The lower bound of $K^{**}$ is taken so that the estimator variance at the lower
bound is equal to the variance obtained whilst considering all the positive observations ($n$
observations) plus a tolerance factor equal to $0.02x(-0.5/\rho)$. The higher bound is set so that
it varies according to the relative lower bound level. The variance estimates of the $\rho$
estimators are computed according to the form given by Fraga Alves et al (2003). Finally,
we reiterate the process to evaluate a new region $K^{***}$, using the estimation $\rho$ & $\beta$ that we
obtained previously with the region $K^{**}$.
Algorithm 2:

1. Given a sample of \( n \) positive observations \((X_1, X_2, ..., X_n)\), evaluate the estimates \( \hat{\lambda}_\tau(k) \) for 36 potential values from -2 to 1.5 as candidates for \( \tau \).

2. Consider \( \hat{\rho}_\tau(k) \), for integer values \( k \in K = \left\{ \frac{2^{0.099}}{n}, \frac{3^{0.099}}{n} \right\} \) and compute their median, denoted, \( \hat{\rho}_\tau \). Next choose the tuning parameter \( \tau^* \approx \arg\min_{k} \sum_{k \in K} \left| \hat{\lambda}_\tau(k) - \hat{\rho}_\tau \right|^2 \).

3. Work then with \( \hat{\rho}_\tau = \hat{\rho}_\tau(k) \) and \( \hat{\rho}_\tau = \hat{\rho}_\tau(k) \), where \( k_1 = \left\lfloor 0.099 \right\rfloor \).

4. Compute
   \[
   \hat{v}_\text{ur}_\tau(k) = \left[2\hat{\lambda}_\tau^2 - 2\hat{\rho}_\tau + 1\right] - \hat{\rho}_\tau^2 \sqrt{4\hat{\lambda}_\tau \hat{\lambda}_\tau (\text{sample size})(k/\hat{\rho}_\tau)},
   \]

5. Choose \( k_{\text{min}} \) such that \( k_{\text{min}} = \min(\hat{\rho}_\tau, k_{\text{max}}) \) and
   \[
   \hat{v}_\text{ur}_\tau(k_{\text{max}}) = \hat{v}_\text{ur}_\tau(n) + 0.02 \frac{n}{\hat{\rho}_\tau}.
   \]

6. Choose \( k_{\text{max}} \) such that
   \[
   k_{\text{max}} = \min\left(n \frac{0.099}{k_{\text{max}}} \right), n \frac{0.099}{n - n^{0.099}}, n - n^{0.099}, n - n^{0.099}
   \]
   and
   \[
   k_{\text{max}} = n - n^{0.099} - n^{0.999} \left(n^{0.999} - n^{0.999}ight).
   \]

7. Consider \( \hat{\rho}_\tau(k) \), for integer values \( k \in K = \left\lfloor k_{\text{max}} \right\rfloor \) and compute their median, denoted \( \hat{\rho}_\tau \). Next choose the tuning parameter \( \tau^{**} \approx \arg\min_{k} \sum_{k \in K} \left| \hat{\lambda}_\tau(k) - \hat{\rho}_\tau \right|^2 \).

8. Work then with \( \hat{\rho}_\tau = \hat{\rho}_\tau(k_{\text{max}}^*) \) and \( \hat{\rho}_\tau = \hat{\rho}_\tau(k_{\text{min}}^*) \), where \( k_1 = \left\lfloor 0.099 \right\rfloor \).

9. Reiterate the steps 4, 5, and 6 with \( \hat{\rho}_\tau \), to evaluate new value of \( k_{\text{min}} \) and \( k_{\text{max}} \).

10. Consider \( \hat{\rho}_\tau(k) \), for integer values \( k \in K = \left\lfloor k_{\text{max}} \right\rfloor \) and compute their median, denoted \( \hat{\rho}_\tau \). Next choose the tuning parameter \( \tau^{***} \approx \arg\min_{k} \sum_{k \in K} \left| \hat{\lambda}_\tau(k) - \hat{\rho}_\tau \right|^2 \).

11. Work then with \( \hat{\rho}_\tau = \hat{\rho}_\tau(k_{\text{max}}^*) \) and \( \hat{\rho}_\tau = \hat{\rho}_\tau(k_{\text{min}}^*) \), where \( k_1 = \left\lfloor 0.099 \right\rfloor \).
3.2.2 **Monte Carlo simulation of the 2 algorithms**

We implemented Monte Carlo simulations of 200 samples of size \( n = 5000 \) from Burr, Student and Frechet distributions and estimated the parameter \( \rho \) for each sample using the algorithm proposed by Gomes, Rodrigues, Pereira and Pestana (alg1) and the algorithm that we suggest in this paper (alg2). We compared the two algorithms by computing the relative standard error of the two estimations.

We show below our findings for

- Burr and Frechet distributions with a EVI equal to 0.5 and different values of \( \rho \) for the Burr distribution (for Frechet, \( \rho \) can take only the value -1).
- Student distributions with 1, 2 and 4 degrees of freedom for which \( \rho \) takes respectively the values of -2, -1 and -0.5.

Similar results were found for the Burr and Frechet distributions with other EVI values. Indeed, we found that the estimation of the parameter \( \rho \) is not sensitive to changes in the value of the EVI.

![Graph 13 - Relative standard error of \( \rho \) estimation (continuous lines = algorithm 1 / dashed lines = algorithm 2) in function of the true absolute value of \( \rho \)](image)

Our algorithm leads to more consistent results and exhibits, in most cases, a lower standard error than the algorithm proposed by Gomes, Rodrigues, Pereira and Pestana (except for Frechet). It outperforms Algorithm 1 when \( \rho \) has small absolute values which is in fact the situation where a good estimation of \( \rho \) is the most required (peak shape of the standard error of the standard estimators and high estimation risk).

To illustrate the consequence of a better performance of the estimation of \( \rho \), we performed Monte Carlo simulations of the relative performance of an estimator of the
Extreme Value Index where $\rho$ estimate needs to be imputed.

The estimator considered is a Minimum Variance Reduced Bias (MVRB) estimator that we will introduce in the next section. We generated 200 samples of size $n=5000$ from Burr, Student and Frechet distributions with various combinations of EVI and $\rho$ values and compared the standard errors of the EVI estimations obtained with our algorithm versus the algorithm proposed by Gomes, Rodrigues, Pereira and Pestana assuming that the threshold is chosen in an optimal way.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|}
\hline
\textbf{EVI values} & 0.25 & 0.5 & 0.75 & 1 & 1.25 & 1.5 \\
\hline
\textbf{Rho values} & -0.25 & 0.61 & 0.64 & 0.66 & 0.61 & 0.61 \\
 & -0.5 & 0.67 & 0.64 & 0.69 & 0.66 & 0.64 & 0.66 \\
 & -0.75 & 1.25 & 1.35 & 1.38 & 1.3 & 1.43 & 1.24 \\
 & -1 & 0.82 & 0.9 & 0.97 & 0.9 & 0.99 & 0.98 \\
 & -1.25 & 0.77 & 0.7 & 0.8 & 0.75 & 0.79 & 0.82 \\
 & -1.5 & 0.77 & 0.81 & 0.73 & 0.77 & 0.96 & 0.96 \\
 & -1.75 & 0.82 & 0.86 & 0.9 & 0.96 & 0.82 & 0.91 \\
 & -2 & 0.91 & 0.9 & 0.81 & 0.9 & 0.98 & 1.01 \\
\hline
\end{tabular}
\end{table}

Table 4 – Comparison of the simulated optimal RSE for the estimation of the EVI using the MVRB technique and both algorithms for the estimation of $\rho$

Ratio = RSE (MVRB with Algorithm 2) / RSE (MVRB with Algorithm 1)

3.2.3 Estimation of the scale second order parameter $\beta$

The estimator of $\beta$ commonly considered is an estimator suggested by Gomes and Martins (2002), with the functional expression

$$
\hat{\beta}_\rho(k) = \left( \frac{k}{n} \right)^{\rho} \sqrt{ \frac{1}{k} \sum_{i=1}^{k} \left( \frac{i}{k} \right)^{\rho} \left[ N_{n}(1) - N_{n}(1-\hat{\rho})(k) \right] } \
$$

with $N_{n}(\alpha)(k) = \frac{1}{k} \sum_{i=1}^{k} \left( \frac{i}{k} \right)^{\alpha} \sum_{i=1}^{\alpha-1} \left( \log X_{n-i+1} - \log X_{n-i} \right)$

and $X_{1}^{*} \leq X_{2}^{*} \leq \ldots \leq X_{n}^{*}$ the order statistics

As for the estimation of the parameter $\rho$, we implemented Monte Carlo simulations of 200 samples of size $n=5000$ from Burr, Student and Frechet distributions and estimated
the parameter $\beta$ for each sample using the $\rho$ estimates performed previously with the two algorithms.

Graph 14 - Relative standard error of $\beta$ estimation (continuous lines = algorithm 1 / dashed lines = algorithm 2) in function of the true absolute value of $\rho$

In this experiment, we can notice that the estimation of the parameter $\beta$ is not very sensitive to the algorithm used for the estimation of $\rho$ but very sensitive to the model underlying the data. The estimator exhibits large bias for Frechet distributions and significant bias for Student distributions which means that some improvement may be still welcome.

We tested another technique to estimate $\beta$ suggested by Gomes, Rodrigues, Perreira and Pestana (2008) based on the log excesses without noticing any improvement. The first technique seems to converge to the second one around the true value at moderate $k$ values but the two techniques have relevant bias at the level $k$ suggested for the estimation ($k=n^{0.995}$) when data are modelled from Frechet and Student distributions.

3.3 Reduced Bias estimators

We present here an attractive approach, which is not part of the standard EVT toolkit: the reduced bias techniques.

The key idea is to deal with the bias term of the Hill estimator in order to:

- make possible the use of larger $k$ to obtain a reduction of the asymptotic variances of the parameter estimates (reduction of the estimation risk) and
- have sample paths with larger regions of stability around the true value to make the choice of $k$ easier (mitigation of the system risk)
We found such techniques very powerful in some cases. The choice to implement them in a company may present the opportunity to get a better model of large losses than its competitors or clients. However, we would like to also highlight that their implementation also presents some risks. Risk to be exposed to:

- unknown practical limitations: incomplete identification of limitations since the techniques were recently developed and
- uncontrolled misspecification risk: misunderstanding of application conditions or lack of tools to assess them.

Indeed, those techniques require the estimation of second order parameters which can be challenging as we saw earlier. They also present misspecification risk since most of them make assumptions on the second order expansion of the distribution underlying the data (in most cases, Hall-Welsh class) in order to get information on the distributional behaviour of the Hill estimator.

We can find in the literature three kinds of reduced bias estimators of the Extreme Value Index.

- The first kind estimates the second order parameters ($\rho$ and $\beta$) together with the EVI parameter $\gamma$ using Maximum Likelihood estimation. Such reduced bias estimators have a strong variance.
- The second kind uses an external estimation of $\rho$ at a higher level $k_1$ than the level $k$ used for $\gamma$. Their estimation variance is reduced but can still be high when $\rho$ has small absolute values.
- The third and most recent kind performs an external estimation of both $\rho$ and $\beta$ and is called the Minimum Variance Reduced Bias (MVRB) estimation. The external estimation of $\rho$ and $\beta$ will maintain the estimation variance at the same level as Hill’s variance on condition that the external estimations of two second order parameters are consistent. However, such a result is difficult to reach in practice since the estimation, as we saw earlier, can be challenging and improvements are still welcome.

### 3.3.1 Minimum Variance Reduced Bias estimator (MVRB)

We present here three MVRB estimators, which have been built for the Hall-Welsh class:

- The first one is the simplest one and was suggested by Caeiro, Gomes and Pestana (2005).
- The second one uses an estimator of \( \gamma / (1 - \rho) \) and was introduced by Beirlant, Figueiredo, Gomes and Vandewalle (2006).
- The third one is a weighted Hill estimator put forward by Gomes, de Haan and Rodrigues (2008).

In the following, we will test only the second one since it appeared to us to be the most powerful.

**EVI estimator**

1. \( MVRB_{Beirlant2005, \beta, \rho, \lambda, \alpha} = \text{Hill}_{X_{\beta, \rho, \lambda, \alpha}} \left( 1 - \frac{\beta}{1 - \rho} \left( \frac{n}{k} \right)^\rho \right) \)

2. \( MVRB_{Beirlant2006, \beta, \rho, \lambda, \alpha} = \text{Hill}_{X_{\beta, \rho, \lambda, \alpha}} \left( 1 - \frac{\beta}{1 - \rho} \left( \frac{n}{k} \right)^\rho \right) D_{\beta, \lambda, \rho} \)

with \( D_{\beta, \lambda, \rho} = \frac{1}{k} \sum_{i=1}^k \left( \frac{i}{k+1} \right)^\rho \left( \log X_{i+1} - \log X_i \right) \)

3. \( MVRB_{Rodrigues2008, \beta, \rho, \lambda, \alpha} = \frac{1}{k} \sum_{i=1}^k \exp \left( - \beta \left( \frac{i}{k} \right)^\rho \frac{\psi \left( \frac{i}{k} \right) \log X_{i+1} - \log X_i}{\rho \ln k} \right) \)

with \( \psi \left( \frac{i}{k} \right) = - \frac{u^{\rho} - 1}{\rho \ln u} \) and with \( X_1 \leq X_2 \leq \ldots \leq X_k \) the order statistics

**VaR estimator**

MVRB techniques require the construction of specific derived quantile estimators. Indeed, even if the EVI is estimated in an unbiased way, tail probability estimators may still exhibit asymptotic bias if based upon the Pareto Distribution approximation. Thus, the derived quantile estimator should not be computed according to the Weissman form (as for the Hill estimator) but according to a bias-corrected Weissman estimator which can take the following form:

\[
\text{Weissman}_{\beta, \rho, \lambda, \alpha}^{(p)} = X_{\rho + \epsilon_x}^{(p)} \quad \text{with} \quad \epsilon_x = \frac{k}{np}
\]

\[
\text{Bias-corrected Weissman}_{\beta, \rho, \lambda, \alpha}^{(p)} = X_{\rho + \epsilon_x}^{(p)} \exp \left( MVRB_{\beta, \rho, \lambda, \alpha} \frac{\rho}{k} \epsilon_x^{\rho - 1} \right)
\]
3.3.2 **Extended Pareto Distribution (EPD)**

We chose also to present one reduced bias estimator of the second kind (only \( \rho \) is estimated externally) suggested by Beirlant, Joossens and Segers (2009) based on an Extended Pareto Distribution (EPD) with the following form:

\[
G_{\gamma, \delta, \rho}(y) = \begin{cases} 
1 - \left[ \frac{1}{\rho} \left( 1 + \delta \left( y^{\rho/\gamma} - 1 \right) \right) \right]^{-1/\gamma}, & \text{if } y > 1, \quad \text{with } \rho < 0 < \gamma \text{ and } \delta > \max(-1, \gamma \rho) \\
0, & \text{if } y \leq 1
\end{cases}
\]

If we take \( \delta = 0 \), we can notice that the ordinary Pareto Distribution (PD) is a member of the EPD family.

Unlike for the MVRB estimator, the distribution can be used to estimate any tail-related risk measure.

The technique works only in the case of a heavy tailed distribution and has been developed for models satisfying a second order condition a little broader than the Hall class.

\[
\overline{F}(x) = C x^{-1/\gamma} \left( 1 + \gamma^{-1} \delta(x) \right) \quad \text{with } C > 0
\]

and \( \delta \) is regularly varying with index \( \rho / \gamma (< 0) \) at infinity.

Models such as a Paretian logarithmic tail such as the Log-gamma distribution are still excluded.

The parameter \( \rho \) still needs to be estimated externally, the same way as for the MVRB estimator. The estimation of other parameters \( \delta \) and \( \gamma \) is carried out using the
Maximum Likelihood technique and imputing estimators of $\rho$:

$$
EPD_{Beirlant2006,\hat{\rho},k,n} = \text{Hill}_{k,n} - \hat{\delta}_{\hat{\rho},k,n} \, \frac{\hat{\rho}}{1 - \hat{\rho}}
$$

with $\hat{\delta}_{\hat{\rho},k,n} = \text{Hill}_{k,n}(1 - 2\hat{\rho})(1 - \hat{\rho})^2 \hat{\rho}^{-1} \left( E_{\hat{\rho}}(\hat{\rho} / \text{Hill}_{k,n}) - \frac{1}{1 - \hat{\rho}} \right)$ and

$$
E_{\hat{\rho},k}(s) = \frac{1}{k} \sum_{i=1}^{k} \left( \frac{X_{\cdot i+1}^*}{X_{\cdot max}^*} \right)^s
$$

The main limitation of the technique is that its variance can be particularly strong for small $|\rho|$ values

$$
\text{Var}(EPD_{Beirlant2006,\hat{\rho},k,n}) = \left( \frac{1 - 2\rho}{\rho} \right)^2 / k
$$

Therefore, the EPD estimator only performs best when MVRB and Hill estimations already perform well. As a consequence, we found that this estimator has a limited contribution to the estimation risk reduction even if it can in some cases improve the estimation obtained with the MVRB estimator.

3.3.3 Monte Carlo simulation of MVRB and EPD techniques

We test here the MVRB suggested by Beirlant, Figueiredo, Gomes and Vandewalle (2006) and the EPD technique along with the Hill, Moment and GPD techniques, for the estimation of the Extreme Value Index and the VaR at a confidence level of 99.9% in the case of two samples of size $n=5000$ from two Burr distributions.

- The first sample has the set of true parameters: EVI = 0.5 & $\rho$ = 0.5 (set1)
- The second sample has the set of true parameters: EVI = 0.3 & $\rho$ = -2 (set2)
EVI estimation

Graph 16 – Hill, Moment, GPD, MVRB and EPD's EVI estimates paths for 2 data samples of size $n=5000$ from 2 different Burr distributions (set1 and set2)

For the 2 sets, MVRB-based estimates of the EVI parameter have large regions of stability around the true value (which makes the choice of the threshold easier) while
- GPD-based estimates behave well only when $\rho=-0.5$ (since $\text{EVI}+\rho=0$) and
- Hill and EPD-based estimates behave well when $\rho=-2$.

VaR estimation

Graph 17 – Hill, Moment, GPD, MVRB and EPD's VaR(99.9%) estimates paths for 2 data samples of size $n=5000$ from 2 different Burr distributions (set1 and set2)
For the 2 sets, MVRB-based estimations of the VaR(99.9%) are still good but less noticeable. This result can be explained by the fact that:

- the bias of the GPD-based EVI estimator is compensated by the second parameter of the fitted GPD
- the MVRB-based VaR estimates are not directly deduced from a fitted distribution but require the construction of a bias-corrected Weissman estimator

In order to compare the standard errors of the estimators, we implemented Monte Carlo simulations of 200 samples of size n = 5000 from Burr, Frechet and Student distributions. We implemented for each technique the estimation of Extreme Value Index (EVI) and the VaR(99.9%) assuming that the optimal threshold is known.

<table>
<thead>
<tr>
<th>Burr parent</th>
<th>EVI values</th>
<th>0.25</th>
<th>0.5</th>
<th>0.75</th>
<th>1</th>
<th>1.25</th>
<th>1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.25</td>
<td>GPD</td>
<td>MVRB</td>
<td>MVRB</td>
<td>MVRB</td>
<td>MVRB</td>
<td>MVRB</td>
<td></td>
</tr>
<tr>
<td>-0.5</td>
<td>MVRB</td>
<td>GPD</td>
<td>MVRB</td>
<td>GPD</td>
<td>MVRB</td>
<td>MVRB</td>
<td></td>
</tr>
<tr>
<td>-0.75</td>
<td>MVRB</td>
<td>MVRB</td>
<td>MVRB</td>
<td>MVRB</td>
<td>MVRB</td>
<td>MVRB</td>
<td></td>
</tr>
<tr>
<td>-1</td>
<td>MVRB</td>
<td>MVRB</td>
<td>MVRB</td>
<td>GPD</td>
<td>MVRB</td>
<td>MVRB</td>
<td></td>
</tr>
<tr>
<td>-1.25</td>
<td>MVRB</td>
<td>MVRB</td>
<td>MVRB</td>
<td>GPD</td>
<td>MVRB</td>
<td>MVRB</td>
<td></td>
</tr>
<tr>
<td>-1.75</td>
<td>MVRB</td>
<td>MVRB</td>
<td>MVRB</td>
<td>MVRB</td>
<td>MVRB</td>
<td>GPD</td>
<td></td>
</tr>
<tr>
<td>-2</td>
<td>HILL</td>
<td>HILL</td>
<td>HILL</td>
<td>HILL</td>
<td>HILL</td>
<td>HILL</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Student parent</th>
<th>Degrees of freedom (v)</th>
<th>4</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>EVI values</td>
<td>0.25</td>
<td>0.5</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Rho values</td>
<td>-0.5</td>
<td>-1</td>
<td>-2</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Frechet parent</th>
<th>EVI values (EVI=1/shape)</th>
<th>0.5</th>
<th>1</th>
<th>1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moment</td>
<td>GPD</td>
<td>GPD</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5 – EVT technique for the estimation of EVI with the lowest RSE according to Monte Carlo simulations of samples of size n = 5000 from models in Hall-Welsh’s class

In this simulation, the MVRB estimator exhibits a good performance.

- In the case of Burr distributions, it is only outperformed by the GPD technique when data follow strictly a GPD (or a Burr with $\gamma + \rho = 0$), and slightly outperformed by Hill when the rate of convergence is already very good ($\rho = -2$).
- In the case of the Frechet and Student distributions, we noted earlier a relevant bias in the $\beta$ estimation (when $\beta \neq 1$) which prevents the estimator from performing as well as for the Burr distribution.
Table 6 – EVT technique for the estimation of VaR(99.9%) with the lowest RSE according to Monte Carlo simulations of samples of size n=5000 from models in Hall-Welsh’s class

<table>
<thead>
<tr>
<th>Burr parent</th>
<th>EVI values</th>
<th>0.25</th>
<th>0.5</th>
<th>0.75</th>
<th>1</th>
<th>1.25</th>
<th>1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rho values</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.25</td>
<td>GPD</td>
<td>GPD</td>
<td>Emp</td>
<td>MVRB</td>
<td>MVRB</td>
<td>MVRB</td>
<td>MVRB</td>
</tr>
<tr>
<td>-0.5</td>
<td>GPD</td>
<td>GPD</td>
<td>MVRB</td>
<td>MVRB</td>
<td>MVRB</td>
<td>MVRB</td>
<td>MVRB</td>
</tr>
<tr>
<td>-0.75</td>
<td>EPD</td>
<td>EPD</td>
<td>GPD</td>
<td>MVRB</td>
<td>MVRB</td>
<td>MVRB</td>
<td>MVRB</td>
</tr>
<tr>
<td>-1</td>
<td>EPD</td>
<td>EPD</td>
<td>GPD</td>
<td>GPD</td>
<td>MVRB</td>
<td>MVRB</td>
<td>MVRB</td>
</tr>
<tr>
<td>-1.25</td>
<td>EPD</td>
<td>EPD</td>
<td>GPD</td>
<td>GPD</td>
<td>GPD</td>
<td>MVRB</td>
<td>MVRB</td>
</tr>
<tr>
<td>-1.5</td>
<td>EPD</td>
<td>GPD</td>
<td>GPD</td>
<td>GPD</td>
<td>GPD</td>
<td>GPD</td>
<td>MVRB</td>
</tr>
<tr>
<td>-1.75</td>
<td>MVRB</td>
<td>EPD</td>
<td>EPD</td>
<td>EPD</td>
<td>GPD</td>
<td>GPD</td>
<td>GPD</td>
</tr>
<tr>
<td>-2</td>
<td>HILL</td>
<td>EPD</td>
<td>EPD</td>
<td>EPD</td>
<td>EPD</td>
<td>GPD</td>
<td>GPD</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Student parent</th>
<th>Degrees of freedom (v)</th>
<th>4</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>EVI values</td>
<td>0.25 0.5 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rho values</td>
<td>-0.5 -1 -2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPD GPD EPD</td>
<td>1 1.49 1.06</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

RSE: Relative Standard Error

Table 7 – Relative performance of the best technique vs the MVRB estimator for the estimation of VaR (99.9%)

In the case of the VaR(99.9%) estimation, even if the MVRB estimator is outperformed by the EPD or the GPD techniques for a large set of distributions, its relative performance is most of the time very close to the best estimator in the case of Burr distributions and is more consistent than the EPD technique.

We can also notice again that its performance for Frechet and Student distributions could be improved by a better estimation of the scale second order parameter.
### Table 8 – Relative performance of Hill vs the MVRB estimator for the estimation of VaR(99.9%)

One important result, here, is that the estimator outperforms almost systematically the Hill estimator even for Student and Frechet distributions.

### Table 9 – Relative performance of GPD technique vs the MVRB estimator for the estimation of VaR(99.9%)

The comparison with the GPD technique is less easy. Indeed, the GPD estimator is an unbiased estimator for combinations of values of γ and ρ such that γ + ρ = 0 and then outperforms the MVRB estimator around these values.

We should emphasize one important feature of the MVRB technique which is not visible in the previous tables. It provides sample paths with larger regions of stability around the true value than other techniques and its standard error function has a “less peaked” shape which makes the choice of k less problematic.
Graph 18 – Simulated relative bias and standard errors of Hill, GPD, MV B’s VaR(99.9%) estimations obtained via Monte Carlo simulations of 200 samples of size n=5000 from different Burr distributions
3.3.4 Comparison with the penultimate estimation

To test the robustness of the MVRB technique, we conducted Monte Carlo simulations of 400 samples of 500 and 5000 observations, from different Log-gamma distributions for which the estimator has not been developed.

We compare the results with the standard EVT techniques and also the penultimate estimation.
Our findings are a little different from those of Degen and Embrechts since they compared the GPD versus the penultimate estimation for a fixed fraction of data equal to 10% for the GPD and a fraction varying according to the sample size for the penultimate approximation (4% for n=250, 2% for n=500 and 1% for n=1000).

<table>
<thead>
<tr>
<th>N=500</th>
<th>Estimator with the lowest RSE at optimal level</th>
<th>Optimal RSE of the best estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EVI values</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Moment</td>
<td>Penult</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>Moment</td>
</tr>
<tr>
<td></td>
<td>GPD</td>
<td>MVRB</td>
</tr>
<tr>
<td></td>
<td>2.5</td>
<td>GPD</td>
</tr>
</tbody>
</table>

RSE > 25%
10% < RSE < 25%
5% < RSE < 10%
RSE < 5%

**Table 10** – EVT technique for the estimation of VaR(99.9%) with the lowest RSE according to Monte Carlo simulations of samples of size n=500 from Log-gamma models

<table>
<thead>
<tr>
<th>Relative performance of GPD</th>
<th>Relative performance of Hill</th>
</tr>
</thead>
<tbody>
<tr>
<td>EVI values</td>
<td>EVI values</td>
</tr>
<tr>
<td>0.5</td>
<td>2.07</td>
</tr>
<tr>
<td>1.5</td>
<td>1.01</td>
</tr>
<tr>
<td>2.5</td>
<td>1</td>
</tr>
<tr>
<td>3.5</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Relative performance of the penultimate approximation</th>
<th>Relative performance of Moment</th>
</tr>
</thead>
<tbody>
<tr>
<td>EVI values</td>
<td>EVI values</td>
</tr>
<tr>
<td>0.5</td>
<td>1.38</td>
</tr>
<tr>
<td>1.5</td>
<td>1.23</td>
</tr>
<tr>
<td>2.5</td>
<td>1.44</td>
</tr>
<tr>
<td>3.5</td>
<td>1.33</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Relative performance of MVRB</th>
<th>Relative performance of empirical quantile</th>
</tr>
</thead>
<tbody>
<tr>
<td>EVI values</td>
<td>EVI values</td>
</tr>
<tr>
<td>0.5</td>
<td>2.42</td>
</tr>
<tr>
<td>1.5</td>
<td>1.22</td>
</tr>
<tr>
<td>2.5</td>
<td>1.5</td>
</tr>
<tr>
<td>3.5</td>
<td>1.45</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ratio</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2</td>
<td>&lt; Ratio &lt; 2</td>
</tr>
<tr>
<td>1.1</td>
<td>&lt; Ratio &lt; 1.2</td>
</tr>
<tr>
<td>Ratio&lt; 1.1</td>
<td></td>
</tr>
</tbody>
</table>

**Table 11** – Relative performances for the estimation of VaR(99.9%). Ratio = RSE at optimal level of the studied technique / RSE at optimal level of the best technique
For combinations of 3 values of the rate log parameter (1.5, 2.5, 3.5) and 3 values for \( \gamma \) (0.5, 1, 1.5) and a low size of samples (500), the MVRB estimator performs surprisingly relatively well but its absolute performance is poor. For other values, the estimator is systematically outperformed by the three standard EVT techniques.

<table>
<thead>
<tr>
<th>Shape log</th>
<th>EVI values</th>
<th>0.5</th>
<th>1</th>
<th>1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>Moment</td>
<td>GPD</td>
<td>GPD</td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>Moment</td>
<td>GPD</td>
<td>MVRB</td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>GPD</td>
<td>Penultimate</td>
<td>MVRB</td>
<td></td>
</tr>
<tr>
<td>3.5</td>
<td>GPD</td>
<td>Penultimate</td>
<td>Penultimate</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Shape log</th>
<th>EVI values</th>
<th>0.5</th>
<th>1</th>
<th>1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>9%</td>
<td>20%</td>
<td>23%</td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>11%</td>
<td>21%</td>
<td>45%</td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>11%</td>
<td>40%</td>
<td>62%</td>
<td></td>
</tr>
<tr>
<td>3.5</td>
<td>13%</td>
<td>46%</td>
<td>64%</td>
<td></td>
</tr>
</tbody>
</table>

Table 12 – EVT technique for the estimation of VaR(99.9%) with the lowest RSE according to Monte Carlo simulations of samples of size \( n=5000 \) from Log-gamma models

### Relative performance of GPD

<table>
<thead>
<tr>
<th>Shape log</th>
<th>EVI values</th>
<th>0.5</th>
<th>1</th>
<th>1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>1.57</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>1.07</td>
<td>1.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>1</td>
<td>1.06</td>
<td>1.13</td>
<td></td>
</tr>
<tr>
<td>3.5</td>
<td>1</td>
<td>1.08</td>
<td>1.21</td>
<td></td>
</tr>
</tbody>
</table>

### Relative performance of Hill

<table>
<thead>
<tr>
<th>Shape log</th>
<th>EVI values</th>
<th>0.5</th>
<th>1</th>
<th>1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>1.23</td>
<td>1.08</td>
<td>1.53</td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>1.29</td>
<td>1.43</td>
<td>1.11</td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>1.59</td>
<td>1.17</td>
<td>1.24</td>
<td></td>
</tr>
<tr>
<td>3.5</td>
<td>1.71</td>
<td>1.15</td>
<td>1.24</td>
<td></td>
</tr>
</tbody>
</table>

### Relative performance of the penultimate approximation

<table>
<thead>
<tr>
<th>Shape log</th>
<th>EVI values</th>
<th>0.5</th>
<th>1</th>
<th>1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>1.22</td>
<td>1.06</td>
<td>1.44</td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>1.28</td>
<td>1.4</td>
<td>1.13</td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>1.4</td>
<td>1</td>
<td>1.07</td>
<td></td>
</tr>
<tr>
<td>3.5</td>
<td>1.55</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Relative performance of MVRB

<table>
<thead>
<tr>
<th>Shape log</th>
<th>EVI values</th>
<th>0.5</th>
<th>1</th>
<th>1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>2.19</td>
<td>1.84</td>
<td>3.08</td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>1.54</td>
<td>1.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>1.91</td>
<td>1.12</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>3.5</td>
<td>1.92</td>
<td>1.16</td>
<td>1.15</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ratio &gt; 2</th>
<th>1.2 &lt; Ratio &lt; 2</th>
<th>1.1 &lt; Ratio &lt; 1.2</th>
</tr>
</thead>
</table>

Table 13 – Relative performance for the estimation of VaR(99.9%).

*Ratio = RSE at optimal level of the studied technique / RSE at optimal level of the best technique*
In the case of bigger sample size (5000 observations), the relative performance of the MVRB estimator is not improved.

The penultimate estimation appears to be a more appropriate technique. It outperforms almost systematically the Hill estimator. Its comparison with the GPD and Moment techniques is however less straightforward:

- For high thresholds (or small k), the penultimate estimation technique outperforms regularly the standard EVT techniques (results similar to the ones of Degen and Embrechts).
- While for lower thresholds (fraction of data superior to 10%), the GPD and Moment estimations for a VaR(99.9%) lead frequently to better results.

Graph 20 – Relative standard errors of Hill, Moment, GPD, MVRB’s and Penultimate VaR(99.9%) estimations obtained via Monte Carlo simulations of 200 samples of size $n=5000$ from different Log-gamma distributions
4. CONCLUSION

We saw, in this paper, useful references to deal with dependent, shifted, censored and contaminated data. We also highlighted one main concern in EVT: the rate of convergence in its limit theorem. Even if data are independent and identically distributed random variables, we have no guarantee of the quality of the statistical inference. Indeed if the model, underlying the data, has a second order expansion with an extremely slow decay term, the standard EVT toolkit is of little help with finite samples and alternatives should be considered.

An additional challenge in EVT is the system risk. None of the standard estimators performs uniformly best. Hill outperforms for large $|\gamma|$ values and small $\gamma$ values. The GPD-based estimator is unbiased when $\gamma+\rho=0$, while the Moment estimator behaves well when $\gamma$ is around 0.5-0.7. The selection of an appropriate threshold is also problematic. Graphical procedures are often used in practice. One should not trust them fully. We are also exposed to differences in their interpretation.

We saw that assuming that the model belongs to a Hall-Welsh class (GPD, Burr, Frechet, Student), enables us to get useful information on the asymptotic bias of the standard EVT estimators via the estimation of second order parameters. We noted some discrepancies in their estimations. We suggested a new algorithm to reduce some of them. However some improvements are still welcome.

Working in the Hall-Welsh class also allowed us to implement reduced bias techniques. We presented a Minimum Variance Reduced Bias Estimator introduced by Beirlant et al, which outperforms systematically the Hill estimator when data are well-modelled by Hall-Welsh’s distributions (GP, Burr, Frechet, Student). It provided us with sample paths with larger regions of stability around the true value and its standard error function has a “less peaked” shape than the other estimators.

In contrast, we found the reduced bias estimator performance very poor if data are modelled by a Log-gamma distribution which does not belong to the Hall-Welsh class.

By using such a technique, we may significantly reduce the estimation and system risk but we are also exposed to mis specification risk if we are not equipped with a proper tool to test the Hall-Welsh assumption.

Finally, in our study, discussion around the practical application of such techniques is missing. It would be valuable to see some in-depth analysis from a company perspective.
REFERENCES


