SCALING OF HIGH-QUANTILE ESTIMATORS

MATTHIAS DEGEN AND PAUL EMBRECHTS,* *Department of Mathematics, ETH Zurich*

Abstract

Enhanced by the global financial crisis, the discussion about an accurate estimation of regulatory (risk) capital a financial institution needs to hold in order to safeguard against unexpected losses has become highly relevant again. The presence of heavy tails in combination with small sample sizes turns estimation at such extreme quantile levels into an inherently difficult statistical issue. We discuss some of the problems and pitfalls that may arise. In particular, based on the framework of second-order extended regular variation, we compare different high-quantile estimators and propose methods for the improvement of standard methods by focussing on the concept of penultimate approximations.

Keywords: Extreme Value Theory; Peaks Over Threshold; Penultimate Approximation; Power Normalization; Second-Order Extended Regular Variation
AMS 2000 Subject Classification: Primary 60G70

Secondary 62G32

1. Introduction

It is fair to say that the global financial system is going through a deep crisis. Whereas for some time a regulatory framework was put into place to avoid systemic risk, the current problems highlight the total insufficiency of this (so-called) Basel framework. Warnings for this were voiced early on; see for instance Daníelsson et al. [6]. Also the weaknesses of Value-at-Risk (VaR), the risk measure required by the Basel framework, were discussed over and over

* Postal address: Department of Mathematics, Raemistrasse 101, CH-8092 Zurich, Switzerland

again; see for instance Nešlehová et al. [17] and references therein. Nevertheless, it has turned out to be extremely difficult to convince regulators to "think again". As a consequence, and mainly spurred on by the subprime crisis, statisticians are increasingly called upon to single out research themes with considerable practical usefulness. A key example of this is the long-term joint project between the Office of the Comptroller of the Currency (OCC) and the National Institute of Statistical Sciences (NISS) on the topic of "Financial Risk Modeling and Banking Regulation". The current paper is motivated by this research program.

Our starting point is the discussion about the estimation of regulatory (risk) capital a financial institution needs to hold in order to safeguard against unexpected losses. Without going into a full description of financial data—be it Market Risk (MR), Credit Risk (CR) or Operational Risk (OR)—it suffices to know that, according to the current regulatory standards in the banking industry (Basel II framework), risk capital has to be calculated (statistically estimated) using the concept of VaR at very high levels of confidence (for MR usually 99% at a 10-day horizon, for CR and OR 99.9%, for economic capital even 99.97%, all three of them at a 1-year horizon). The credit crisis prompted the introduction of an extra 99.9%, 1-year capital charge for MR, the so-called Incremental Risk Charge; see Basel Committee [20]. Because of the extreme quantile levels required, early on extreme value theory (EVT) was recognized as a potentially useful tool. However, and this often from practice, critical voices have been raised against an imprudent use of standard EVT. In the context of *quantitative risk management (QRM)*, the use of EVT-based high-quantile estimators may indeed be a delicate issue and warrants careful further study.

The aim of our paper is twofold. In a first and more theoretical part, we analyze different choices of normalization and their influence on the rate of convergence in certain limit laws underlying EVT. More precisely, we compare linear and power norming for high-risk scenarios and quantiles which leads to techniques that are not part of the standard EVT toolkit. In particular we propose the use of so-called *penultimate approximations* to estimate extreme quantiles. The idea of penultimate approximations goes back to Fisher and Tippett [9], its potential for practical applications however seems to have received little attention so far; see Degen and Embrechts [7] for some references.

In a second part, concrete applications of the methodology developed in the first part are discussed. We compare the performance of different high-quantile estimators. One method increasingly championed in practice estimates quantiles at a lower level (e.g. 99%) and then scales up to the desired higher level (e.g. 99.9%) according to some scaling procedure to be specified. In this context, the usefulness of penultimate approximations in situations of very heavy tails together with small sample sizes (typical for OR) is highlighted.

2. Univariate EVT

We assume the reader to be familiar with univariate EVT, as presented for instance in Embrechts et al. [8] or in de Haan and Ferreira [12]. Throughout we assume that our loss data $X > 0$ are modeled by a continuous distribution function (df) F with upper end-point $x_F \leq \infty$ and standardly write $\overline{F} = 1 - F$. The corresponding tail quantile function is denoted by $U(t) = F^{\leftarrow}(1-1/t)$, where $F^{\leftarrow}$ denotes the (generalized) inverse of F. To avoid confusion we will—where necessary—denote the df and the tail quantile function of a random variable (rv) X by F_X and U_X, respectively.

As our focus is on the application of EVT-based methods to quantitative risk management, we prefer to work within the framework of exceedances (*Peaks Over Threshold (POT)* method) rather than within the classical framework of *block-maxima*. The two concepts however are closely linked as the next result shows; see de Haan and Ferreira [12], Theorem 1.1.6.

Proposition 2.1. *For $\xi \in \mathbb{R}$ the following are equivalent.*

i) There exist constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that

$$\lim_{n\to\infty} F^n\left(a_n x + b_n\right) = H_\xi(x) = \exp\left\{-\left(1+\xi x\right)^{-1/\xi}\right\}, \tag{1}$$

for all x with $1 + \xi x > 0$.

ii) There exists a measurable function $a(.) > 0$ *such that for* $x > 0$,

$$\lim_{t \to \infty} \frac{U(tx) - U(t)}{a(t)} = D_\xi(x) = \frac{x^\xi - 1}{\xi}. \tag{2}$$

iii) There exists a measurable function $f(.) > 0$ *such that*

$$\lim_{t \to x_F} \frac{\overline{F}\,(t + xf(t))}{\overline{F}\,(t)} = (1 + \xi x)^{-1/\xi}\,, \tag{3}$$

for all x *for which* $1 + \xi x > 0$.

Moreover, (1) holds with $b_n = U(n)$ *and* $a_n = a(n)$*. Also, (3) holds with* $f(t) = a\left(1/\overline{F}(t)\right)$*.*

Definition 2.1. A df F satisfying (1) is said to belong to the *linear maximum (l-max) domain of attraction* of the extreme value distribution H_ξ and we write $F \in D_l^{max}(H_\xi)$. For necessary and sufficient conditions for distributions F to belong to $D_l^{max}(H_\xi)$ we refer to de Haan and Ferreira [12], Chapter 1.

Domain of attraction conditions have been formulated directly in terms of regular variation of $\overline{F}$ at $x_F \leq \infty$ for the cases $\xi > 0$ and $\xi < 0$, but not for the case $\xi = 0$; see Gnedenko [10]. The novelty of Proposition 2.1 (originally due to de Haan [11]) is that it treats the domain of attraction conditions for the three cases in a unified way by making use of the more general concept of *extended regular variation (ERV)* for U. Recall that a function U is said to be of extended regular variation with index $\xi \in \mathbb{R}$ and with auxiliary function $a(.)$ if it satisfies (2); see de Haan and Ferreira [12], Appendix B.2. In that case we write $U \in ERV_\xi(a)$.

Remark 2.1. Even within the unified framework of ERV, the case $\xi = 0$ is still somewhat special. Acting as limiting cases, the right hand sides in (2) and (3) are interpreted as $\log x$ and e^{-x} respectively. In that case, U and $1/\overline{F}$ are said to be of Π-variation and Γ-variation, respectively, and we write $U \in \Pi(a)$ (or $U \in ERV_0$) and $1/\overline{F} \in \Gamma(f)$.

From a theoretical point of view, this full generality of the framework of extended regular variation is certainly to be appreciated. For applications to QRM however, a framework treating $\xi \geq 0$ but not $\xi < 0$ in an as simple as possible way is to be preferred. This is done below basically by working with $\log U$ instead of U.

3. Asymptotic properties of normalized high-risk scenarios and quantiles

For a positive rv $X \sim F$ we introduce the notation of X^t, which is defined as the rv X, conditioned to exceed the threshold $t > 0$. Within QRM, X^t is often referred to as a *high-risk scenario*; see also Balkema and Embrechts [1] for this terminology.

With this notation, Proposition 2.1 iii) states that high-risk scenarios, linearly normalized, converge weakly to a non-degenerate limit, i.e. for $\xi \in \mathbb{R}$ and $x > 0$,

$$P\left(\frac{X^t - t}{f(t)} > x\right) = \frac{\overline{F}(t + xf(t))}{\overline{F}(t)} \to -\log H_\xi(x) = (1 + \xi x)^{-1/\xi}, \quad t \to x_F, \tag{4}$$

for some measurable function $f(.) > 0$. In that case we shall say that F belongs to the *linear POT (l-POT) domain of attraction* of H_ξ and write $F \in D_l^{POT}(H_\xi)$.

While the limit behavior of random variables (exceedances as well as block-maxima) under linear normalizations is well understood and frequently used in applications, the theory under non-linear normalizations has been studied less. Pantcheva [18] and Mohan and Ravi [15] developed a theory of power norming within the block-maxima framework.

We shall adopt this idea of non-linear norming and study the limit behavior of power normalized high-risk scenarios. Inspired by Barakat et al. [2], who compare the convergence rates under linear and power normalization within the block-maxima setting, we study the first- and second-order asymptotic behavior of power-normalized high-risk scenarios and quantiles.

Definition 3.1. We say that a df F belongs to the *power POT (p-POT) domain of attraction* of some non-degenerate df K and write $F \in D_p^{POT}(K)$, if there exists a measurable function $g(.) > 0$ such that the (power) normalized high-risk scenario $(X^t/t)^{1/g(t)}$ converges weakly to K, in the sense that

$$P\left(\left(X^t/t\right)^{1/g(t)} > x\right) \to \overline{K}(x), \quad t \to x_F, \tag{5}$$

for every continuity point $x > 0$ of K.

Introducing logarithms proves useful at this point, as it provides a link between the two concepts of linear and power norming for high-risk scenarios. In particular we have the following result about the relation between the respective domains of attraction D_l^{POT} and D_p^{POT}.

Proposition 3.1. *For $X > 0$ with df F_X and for $\xi \in \mathbb{R}$ the following holds:*

$$
\begin{aligned}
&i) \quad F_{\log X} \in D_l^{POT}(H_\xi) \quad \Longleftrightarrow \quad F_X \in D_p^{POT}(K_\xi),\\
&ii) \quad F_X \in D_l^{POT}(H_\xi) \qquad \Longrightarrow \quad F_X \in D_p^{POT}(K_{\xi_-}),
\end{aligned}
$$

where $\overline{K}_\xi(x) = -\log H_\xi(\log x)$ for $x > 0$ and $\xi_- = \xi \wedge 0$.

Proof. *i)* Let $\xi \in \mathbb{R}$ and $x > 0$. Setting $Y = \log X$, the corresponding high-risk scenario satisfies $Y^s = \log(X^t)$ for $s = \log t$ and thus it immediately follows that

$$\lim_{s \to x_F} P\left(\frac{Y^s - s}{f(s)} > x\right) = -\log H_\xi(x) \quad \Longleftrightarrow \quad \lim_{t \to x_F} P\left(\left(\frac{X^t}{t}\right)^{1/g(t)} > x\right) = \overline{K}_\xi(x),$$

where $f(s) = g(t)$ and with $s = \log t$.

ii) Let $x > 0$ and assume $F \in D_l^{POT}(H_\xi)$, i.e. for some $f(.) > 0$ and with $x > 0$,

$$P\left(\frac{X^t - t}{f(t)} > x\right) \to (1 + \xi x)^{-1/\xi}, \quad t \to x_F.$$

We make use of the fact that the convergence above is uniformly in t. Moreover, define $\lambda_t(x) = \frac{tx^{g(t)} - t}{f(t)}$ and observe that if $\lim_{t \to x_F} \lambda_t(x) =: \lambda_\infty(x)$ exists, we have for every $x > 0$,

$$P\left(\left(\frac{X^t}{t}\right)^{1/g(t)} > x\right) = \frac{\overline{F}(tx^{g(t)})}{\overline{F}(t)} = \frac{\overline{F}\left(t + \lambda_t(x) f(t)\right)}{\overline{F}(t)} \to (1 + \xi \lambda_\infty(x))^{-1/\xi}, \quad t \to x_F,$$

Now, set $g(t) = f(t)/t$ so that $\lambda_t(x) = \dfrac{x^{f(t)/t} - 1}{f(t)/t}$ for $x > 0$.

$\xi > 0$: In this case $g(t) \to \xi$, as $t \to x_F$; see de Haan and Ferreira, Theorem 1.2.5. Therefore, the limit λ_∞ exists, is finite and we have $\lim_{t \to x_F} \lambda_t(x) = (x^\xi - 1)/\xi$.

$\xi < 0$: Note first that $x_F < \infty$. Moreover we have $f(t)/(x_F - t) \to -\xi$ as $t \to x_F$ (see de Haan and Ferreira [12], Theorem 1.2.5.) and hence $g(t) \to 0$ for $t \to x_F$. Therefore we obtain $\lim_{t \to x_F} \lambda_t(x) = \log x$.

$\xi = 0$: For $\xi = 0$, the right endpoint x_F may be finite or infinite. Moreover, $f(.)$ is asymptotically equivalent to a function $\widetilde{f}(.)$, whose derivative vanishes at x_F. For the case $x_F = \infty$ we thus have

$$\frac{\widetilde{f}(t) - \widetilde{f}(t_0)}{t} = \frac{1}{t}\int_{t_0}^{t} \widetilde{f}'(s)ds \to 0, \quad t \to \infty.$$

Therefore $\widetilde{f}(t)/t \to 0$ as $t \to x_F$ (and hence also $g(t) \to 0$), which in turn implies $\lim_{t \to x_F} \lambda_t(x) = \log x$.

In the case $x_F < \infty$, $\widetilde{f}(t) \to 0$ as $t \to x_F$ (and hence also $g(t) \to 0$); see de Haan and Ferreira [12], Theorem 1.2.6. Therefore we obtain $\lim_{t \to x_F} \lambda_t(x) = \log x$.

Altogether, $F \in D_l^{POT}(H_\xi)$ with $\xi \in \mathbb{R}$ implies that for every $x > 0$ and as $t \to x_F$,

$$P\left(\left(\frac{X^t}{t}\right)^{1/g(t)} > x\right) \to \begin{cases} (1+\xi\lambda_\infty(x))^{-1/\xi} & = \begin{cases} x^{-1}, & \xi > 0, \\ (1+\xi\log x)^{-1/\xi}, & \xi < 0, \end{cases} \\ e^{-\lambda_\infty(x)} & = \quad x^{-1}, \qquad\qquad \xi = 0, \end{cases}$$

i.e. $F \in D_p^{POT}(K_{\xi_-})$, where $\xi_- = \xi \wedge 0$. This finishes the proof.

For later purposes we shall reformulate Proposition 3.1 in terms of quantiles. Due to the convergence properties of inverse functions (see Resnick [19], Proposition 0.1) this is immediate and we have the following result.

Corollary 3.1. *For $X > 0$ with tail quantile function U_X and $\xi \in \mathbb{R}$ the following holds:*

$$\begin{aligned} &i) \quad U_{\log X} \in ERV_\xi(a) \iff \log U_X \in ERV_\xi(a), \\ &ii) \quad U_X \in ERV_\xi(a) \implies \log U_X \in ERV_{\xi_-}(b), \end{aligned}$$

where $\xi_- = \xi \wedge 0$ and $b(t) = a(t)/U(t)$ for some measurable function $a(.) > 0$.

Remarks 3.1. 1) According to Assertion *ii)* of Corollary 3.1, convergence of linearly normalized quantiles $U(tx)$, i.e.

$$\frac{U(tx) - U(t)}{a(t)} \to D_\xi(x) = \frac{x^\xi - 1}{\xi}, \quad t \to \infty, \tag{6}$$

for some $x > 0$, implies convergence of *power normalized* quantiles, i.e.

$$\left(\frac{U(tx)}{U(t)}\right)^{1/b(t)} \to \exp\left\{D_{\xi_-}(x)\right\}, \quad t \to \infty. \tag{7}$$

In the case of main interest for QRM applications, i.e. for $\xi \geq 0$, this rewrites as $U \in ERV_\xi(a) \Rightarrow \log U \in \Pi(b)$.

2) The respective converse implications in *ii*) of Proposition 3.1 and Corollary 3.1 do not hold; D_p^{POT} attracts in fact more distributions than D_l^{POT}. Consider for example $\overline{F}_X(x) = (\log x)^{-1}$ with $x > e$, hence $F_X \notin D_l^{POT}$ but $F_X \in D_p^{POT}$.

Note that for $F \in D_p^{POT}(K)$, the possible limit laws K are unique up to what we might call *p-types* (in the POT setting), where we call two dfs K_1 and K_2 of the same *p-type* if $K_1(x) = K_2(x^p)$ for some $p > 0$.

Proposition 3.2. (Convergence to p-types.) *Let $X \sim F$ be a positive rv and assume K_1 and K_2 are two non-degenerate distribution functions.*

i) If there exist measurable functions $g_1(.) > 0$ and $g_2(.) > 0$, such that for $x > 0$

$$\frac{\overline{F}\left(tx^{g_1(t)}\right)}{\overline{F}(t)} \to \overline{K}_1(x), \quad \frac{\overline{F}\left(tx^{g_2(t)}\right)}{\overline{F}(t)} \to \overline{K}_2(x), \quad t \to x_F, \tag{8}$$

then

$$\lim_{t \to x_F} \frac{g_2(t)}{g_1(t)} = p > 0 \tag{9}$$

and

$$K_2(x) = K_1(x^p). \tag{10}$$

ii) If (9) holds, then either of the two relations in (8) implies the other and (10) holds.

Proof. ii) Assume that (9) holds and that $\overline{F}\left(tx^{g_1(t)}\right)/\overline{F}(t) \to \overline{K}_1(x)$ as $t \to x_F$. From the theory of ERV it clear that the existence of a non-degenerate limit K implies that necessarily $K(x) = 1 - (1 + \xi \log x)^{-1/\xi}$. Since the limit laws K are continuous, uniform convergence holds and we obtain

$$\frac{\overline{F}\left(tx^{g_2(t)}\right)}{\overline{F}(t)} = \frac{\overline{F}\left(t\left(x^{g_2(t)/g_1(t)}\right)^{g_1(t)}\right)}{\overline{F}(t)} \to \overline{K}_1(x^p), \quad t \to x_F.$$

i) Assume that the two relations in (8) hold and set $V(t) = F^{\leftarrow}(1-t)$ and $W_i(t) = K_i^{\leftarrow}(1-t)$ for $0 < t < 1$ and $i = 1, 2$. As K_1 and K_2 are non-degenerate, we may find points x_1, x_2 such that $W_1(x_1) > W_1(x_2)$ and $W_2(x_1) > W_2(x_2)$. Due to the convergence properties of generalized inverse functions, we have

$$\lim_{t \to x_F} \left(\frac{V\left(\overline{F}(t)x_i\right)}{t} \right)^{1/g_j(t)} = W_j(x_i), \quad i, j \in \{1, 2\}\,.$$

Taking logarithms we find

$$\frac{1}{g_j(t)} \log \frac{V\left(\overline{F}(t)x_1\right)}{V\left(\overline{F}(t)x_2\right)} \to \log \frac{W_j(x_1)}{W_j(x_2)} > 0, \quad t \to x_F, \quad j \in \{1, 2\}\,.$$

From this we obtain

$$\lim_{t \to x_F} \frac{g_2(t)}{g_1(t)} = \log \frac{W_1(x_1)}{W_1(x_2)} \Big/ \log \frac{W_2(x_1)}{W_2(x_2)} =: p > 0,$$

which finishes the proof.

In order to appreciate the approximations under linear and power norming, (6) and (7) respectively, we need to quantify and compare the goodness of these approximations. More precisely, we are interested in a comparison of the convergence rates for normalized quantiles in (6) and (7)—or equivalently of the rates for corresponding normalized high-risk scenarios in (4) and (5). It turns out that a judicious choice of the power normalization $b(.)$, respectively $g(.)$, may improve the convergence rate over linear norming. This may be important for applications, as we might hope to improve the accuracy of standard EVT-based high-quantile estimators when using power norming.

3.1. Second-order asymptotics of normalized quantiles

In the sequel we prefer to work with quantiles U rather than distribution tails $\overline{F}$. However, any statement formulated in the $U-$framework may equivalently be expressed in the $\overline{F}-$framework. Moreover, while we worked in full generality (i.e. $\xi \in \mathbb{R}$) so far, we shall henceforth restrict ourselves to the case $\xi \geq 0$, most of interest for QRM applications. Similar results for the case $\xi < 0$ may be worked out.

In order to avoid unnecessary technicalities and to exclude pathological cases we shall throughout assume sufficient smoothness for U. For our purposes, the following representation for U turns out to be convenient to work with:

$$U(t) = e^{\varphi(\log t)}, \quad \varphi(t) = \int_1^{e^t} \frac{ds}{u(s)} + c,$$

where $u(s) = U(s)/U'(s)$ and $c = \log U(1)$. Furthermore we shall assume that

(A1) the *von Mises condition* holds, i.e. $tU''(t)/U'(t) \to \xi - 1$, for some $\xi \geq 0$; see de Haan and Ferreira [12] for details.

Assumption (A1) is equivalent to assuming $\varphi' \to \xi \geq 0$ together with $\varphi''/\varphi' \to 0$. It reflects the fact that the log-log plot φ of U is assumed to behave "nicely" in the sense of being ultimately linear, i.e. with converging slope φ' and vanishing convexity φ''. Moreover, (A1) is sufficient to guarantee $U \in ERV_\xi(a)$, i.e. for $x > 0$

$$\frac{U(tx) - U(t)}{a(t)} \to D_\xi(x) = \frac{x^\xi - 1}{\xi}, \quad t \to \infty, \tag{11}$$

for some measurable function $a(.) > 0$. Note that from the theory of extended regular variation it is clear that $a(t)/U(t) \to \xi \geq 0$. By Corollary 3.1, this in turn implies $\log U \in \Pi(b)$ and hence

$$\left(\frac{U(tx)}{U(t)}\right)^{1/b(t)} \to x, \quad t \to \infty, \tag{12}$$

where $b(t) = a(t)/U(t) > 0$ and such that $b(t) \to \xi$.

Our interest is in second-order results for (11) and (12), i.e. we want to consider functions A and B with $\lim_{t\to\infty} A(t) = \lim_{t\to\infty} B(t) = 0$, which for $\xi \geq 0$ and $x > 0$ satisfy

$$\frac{\frac{U(tx)-U(t)}{a(t)} - D_\xi(x)}{A(t)} \to S(x), \quad t \to \infty, \tag{13}$$

and

$$\frac{D_\xi\left(\left(\frac{U(tx)}{U(t)}\right)^{1/b(t)}\right) - D_\xi(x)}{B(t)} \to T(x), \quad t \to \infty, \tag{14}$$

with $b(t) = a(t)/U(t)$ and for some non-trivial limits $S(.)$ and $T(.)$. Again, the limiting case $D_0(.)$ is interpreted as $\log(.)$.

Clearly, the precise conditions for (13) and (14) to hold will depend on the choice of the normalization $a(.)$, respectively $b(.)$. Different choices of normalization may lead to different asymptotics in the respective second-order relations. We discuss the following three cases.

Case I: $b_1(t) \equiv \xi$ for $\xi > 0$

For this choice of power normalization, the limit relations (13) and (14) coincide and we have the following second-order asymptotics.

Proposition 3.3. *Suppose $U(t) = e^{\varphi(\log t)}$ is twice differentiable and let $A_1(t) = ta'(t)/a(t) - \xi$ with $a(t) = \xi U(t)$ for some $\xi > 0$. Assume that (A1) holds, that φ' is ultimately monotone, and that*

(A2) $\lim_{t\to\infty} \varphi''(t)/(\varphi'(t) - \xi) = \rho$, *for some* $\rho \le 0$.

Then, for $x > 0$,

$$\lim_{t\to\infty} \frac{\frac{U(tx)}{U(t)} - x^{\xi}}{A_1(t)} = H_{\xi,\rho}(x) = x^{\xi} D_{\rho}(x),$$

where $D_{\rho}(x) = \frac{x^{\rho}-1}{\rho}$.

Proof. See Degen and Embrechts [7], Theorem 3.1.

Case II: $b_2(t) = tU'(t)/U(t)$ for $\xi \ge 0$

An intuitive reasoning behind this choice of normalization may be given by noting that $b(t) = tU'(t)/U(t) = \varphi'(\log t)$ is the slope of the log-log plot φ of U. In the sequel we will therefore refer to $b(t) = \varphi'(\log t)$ as the *local slope* (or *local tail index*) of the log-log plot of U at points t (as opposed to the *ultimate slope* $\xi = \varphi'(\infty)$ typically considered in standard EVT); see also Degen and Embrechts [7]. We obtain the following second-order asymptotics for linear and power normalized quantiles.

Proposition 3.4. *Suppose $U(t) = e^{\varphi(\log t)}$ is twice differentiable and let $A_2(t) = ta'(t)/a(t) - \xi$ with $a(t) = tU'(t)$. Assume that (A1) holds for some $\xi \ge 0$, that φ'' is ultimately monotone, and* *that* $|A_2| \in RV_\rho$*, for some* $\rho \leq 0$*. Then, for* $x > 0$*,*

$$\lim_{t\to\infty} \frac{\frac{U(tx)-U(t)}{a(t)} - D_\xi(x)}{A_2(t)} = S_{\xi,\rho}(x),$$

where

$$S_{\xi,\rho}(x) = \int_1^x s^{\xi-1}\int_1^s y^{\rho-1}dyds = \begin{cases} \frac{1}{\rho}\left(D_{\xi+\rho}(x) - D_\xi(x)\right), & \rho < 0, \\ \frac{1}{\xi}\left(x^\xi \log x - D_\xi(x)\right), & \rho = 0, \xi > 0, \\ \frac{(\log x)^2}{2}, & \rho = \xi = 0. \end{cases}$$

Proof. See de Haan and Ferreira [12], Theorem 2.3.12.

Proposition 3.5. *Suppose* $U(t) = e^{\varphi(\log t)}$ *is three times differentiable and let* $B_2(t) = tb'(t)/b(t)$ *with* $b(t) = tU'(t)/U(t)$*. Assume that (A1) holds for some* $\xi \geq 0$*, that* φ'' *is ultimately monotone, and that*

(A3) $\lim_{t\to\infty} \varphi'''(t)/\varphi''(t) = \rho$*, for some* $\rho \leq 0$*.*

Then, for $x > 0$*,*

$$\lim_{t\to\infty} \frac{D_\xi\left(\left(\frac{U(tx)}{U(t)}\right)^{1/b(t)}\right) - D_\xi(x)}{B_2(t)} = T_{\xi,\rho}(x),$$

where

$$T_{\xi,\rho}(x) = \begin{cases} \frac{x^\xi}{\rho}\left(D_\rho(x) - \log x\right), & \rho < 0, \\ x^\xi \frac{(\log x)^2}{2}, & \rho = 0. \end{cases}$$

Proof. We rewrite (14) for $t \to \infty$ as

$$\begin{aligned} \frac{D_\xi\left(\left(\frac{U(tx)}{U(t)}\right)^{1/b(t)}\right) - D_\xi(x)}{B(t)} &\sim x^\xi \frac{\log U(tx) - \log U(t) - b(t)\log x}{b(t)B(t)} \\ &= x^\xi \int_1^x \frac{b(ts) - b(t)}{b(t)B(t)}\frac{1}{s}ds. \end{aligned}$$

For the above integral to converge to a non-trivial limit, it is sufficient to require $b \in ERV_\rho$ for some $\rho \leq 0$ and with auxiliary function $b(t)B(t)$. In that case, it is clear from the theory of extended regular variation that, in the case $\rho < 0$, we may take the auxiliary function to satisfy $b(t)B(t) = tb'(t)$, if b' is ultimately monotone.

For the special case $\rho = 0$ note that, when b' is ultimately monotone, $b' \in RV_{-1}$ if and only if $b \in \Pi(c)$. In this case one may choose $c(t) = tb'(t)$. This follows from the Monotone Density Theorem for $\Pi-$Variation; see Bingham et al. [4], Theorem 3.6.8.

Altogether, $(A3)$ being equivalent to $tb''(t)/b'(t) \to \rho - 1$, together with the assumption of φ'' (i.e. b') being ultimately monotone ensures that $b \in ERV_\rho(c)$ for some $\rho \leq 0$ and such that we may choose $c(t) = tb'(t)$. By the Uniform Convergence Theorem for ERV(see Bingham et al. [4], Theorem 3.1.7a), the convergence

$$\lim_{t\to\infty} \frac{b(ts) - b(t)}{tb'(t)} = \begin{cases} \frac{x^\rho - 1}{\rho}, & \rho < 0, \\ \log x, & \rho = 0. \end{cases}$$

holds locally uniformly on $(0, \infty)$ which finishes the proof.

Case III: $b_3(t) = \log U(t) - 1/t \int_{t_0}^t \log U(s)ds$ for some $t_0 > 0$ and for $\xi \geq 0$

Compared with the previous two cases, this choice of $b(.)$ does not seem to be very intuitive. We shall therefore briefly give a heuristic argument about its raison d'être in the literature. Recall from Case II that in smooth cases we may choose $b(.)$ as the *local slope* of the log-log plot of U, i.e. $b(t) = \varphi'(\log t)$. However, if one does not want to a priori assume differentiabilty, φ' need not exist. In that case, by Karamata's Theorem, $\varphi(\log t)$ is of the same order as its average $\widetilde{\varphi}(\log t) := \frac{1}{t}\int_{t_0}^t \varphi(\log s)ds$, for some $0 < t_0 < t$, i.e. $\widetilde{\varphi}(\log t)/\varphi(\log t) \to 1$ as $t \to \infty$. However, unlike φ, $\widetilde{\varphi}$ is always differentiable and hence—similar to the (smooth) Case II—one may choose $b(t) = \widetilde{\varphi}'(\log t) = \varphi(\log t) - \frac{1}{t}\int_{t_0}^t \varphi(\log s)ds$ with $0 < t_0 < t$. Following the terminology in Case II, we might refer to $\widetilde{\varphi}'$ as a kind of *local "pseudo" slope* of φ (or *local "pseudo" tail index*).

Proposition 3.6. *Suppose $U(t) = e^{\varphi(\log t)}$ is twice differentiable and let $B_3(t) = tb'(t)/b(t)$ with $b(t) = \log U(t) - 1/t \int_{t_0}^t \log U(s)ds$ for some $t_0 > 0$. Assume that (A1) holds for some $\xi \geq 0$, that b' is ultimately monotone, and that*

(A4) $\lim_{t\to\infty} \varphi''(\log t)/\left(\varphi'(\log t) - b(t)\right) - 1 = \tau$, *for some* $\tau \leq 0$.

Then, for $x > 0$,

$$\lim_{t\to\infty} \frac{D_\xi\left(\left(\frac{U(tx)}{U(t)}\right)^{1/b(t)}\right) - D_\xi(x)}{B_3(t)} = T_{\xi,\tau}(x) = x^\xi \left(D_\tau(x) + S_{\xi_-,\tau}(x)\right).$$

Proof. Assumption (A4) rewrites as $tb''(t)/b'(t) \to \tau - 1$, as $t \to \infty$, and, together with b' being ultimately monotone, ensures that $b \in ERV_\tau(c)$ with $c(t) = tb'(t)$. For $b(t) = \log U(t) - \frac{1}{t}\int_{t_0}^t \log U(s)ds$ and for some $x > t_0$ we obtain by partial integration

$$\begin{aligned}\int_{t_0}^x \frac{b(t)}{t}dt &= \int_{t_0}^x \frac{\log U(t)}{t}dt - \int_{t_0}^x \int_{t_0}^t \frac{\log U(s)}{t^2}dsdt \\ &= \frac{1}{x}\int_{t_0}^x \log U(s)ds = \log U(x) - b(x),\end{aligned}$$

so that

$$\log U(x) = b(x) + \int_{t_0}^x \frac{b(t)}{t}dt.$$

Therefore, by the Uniform Convergence Theorem for ERV (see Bingham et al. [4], Theorem 3.1.7a), we obtain for $t \to \infty$,

$$\begin{aligned}\frac{D_\xi\left(\left(\frac{U(tx)}{U(t)}\right)^{1/b(t)}\right) - D_\xi(x)}{B_3(t)} &\sim x^\xi \frac{\log U(tx) - \log U(t) - b(t)\log x}{b(t)B_3(t)} \\ &= x^\xi\left(\frac{b(tx) - b(t)}{tb'(t)} + \int_1^x \frac{b(ts) - b(t)}{tb'(t)}\frac{1}{s}ds\right) \\ &\to x^\xi\left(\frac{x^\tau - 1}{\tau} + S_{\xi_-,\tau}(x)\right),\end{aligned}$$

where $S_{\xi,\tau}$ is as in Proposition 3.4.

Concerning a second-order result under linear norming in the Case III, we draw on the work of Vanroelen [22]. The author relates different second-order relations for cases where the normalization $a(.)$ is replaced by $\widetilde{a}(.)$ with $a(t) \sim \widetilde{a}(t)$, as $t \to \infty$. We have the following result.

Proposition 3.7. *Suppose that the assumptions of Proposition 3.4 hold, as well as (A3) for some $\rho \le 0$ and (A4) with $\tau \ne -1$. Define $\widetilde{a}(t) = U(t)\left(\log U(t) - 1/t\int_{t_0}^t \log U(s)ds\right)$ for some $t_0 > 0$. Then, for $x > 0$,*

$$\lim_{t\to\infty} \frac{\frac{U(tx)-U(t)}{\widetilde{a}(t)} - D_\xi(x)}{A_3(t)} = S_{\xi,\rho,\tau}(x),$$

where $A_3(.) = A_2(.)$ *and* $A_2(.)$ *is as in Proposition 3.4 and with*

$$S_{\xi,\rho,\tau}(x) = \begin{cases} \frac{1}{\rho}\left(D_{\xi+\rho}(x) - (1+\rho\tau)\, D_\xi(x)\right), & \rho < 0, \\ \frac{1}{\xi}\left(x^\xi \log x + (1+\xi\tau)\, D_\xi(x)\right), & \rho = 0, \xi > 0, \\ \frac{(\log x)^2}{2} - \tau \log x, & \rho = \xi = 0. \end{cases}$$

Proof. In terms of φ and its derivatives we have $B_2(t) = \varphi''(r)/\varphi'(r)$ and $A_2(t) = B_2(t) + A_1(t) = \varphi''(r)/\varphi'(r) + \varphi'(r) - \xi$, where $r = \log t$ and with A_i and B_i as defined above. Moreover, $B_3(t) = \varphi'(\log t)/b_3(t) - 1$, so that Assumption (A4) may be rewritten as $B_2(t)/B_3(t) - 1 \to \tau$ for $t \to \infty$ and with $\tau \leq 0$.

The proof now follows from Proposition 1.3.1 of Vanroelen [22] by remarking that Assumption (A4) with $\tau \neq -1$ implies Condition (1.17) of Vanroelen [22]. Indeed, with $a(t) = tU'(t)$ as in Proposition 3.4, we have

$$\lim_{t\to\infty} \frac{1}{A_2(t)} \left(\frac{\widetilde{a}(t)}{a(t)} - 1 \right) = \lim_{t\to\infty} \frac{B_3(t)}{A_2(t)} = \lambda, \tag{15}$$

for some $\lambda \in \mathbb{R}$ ($\lambda = 0$ in the case $\rho = 0$).

3.2. Comparison of the converence rates under linear and power norming

It is not surprising that in the case of a constant power normalization (Case I), no improvement may be achieved when using power norming for quantiles $U(tx)$ instead of linear norming. This however need not be the case for non-constant power normalizations. Indeed, a comparison of the respective second-order limit relations in the Cases II and III shows that there are situations in which the convergence rate for power normalized quantiles is faster than when a linear normalization is applied.

Case II: $b_2(t) = tU'(t)/U(t)$

Recall from the proof of Proposition 3.7 that we have $B_2(t)/A_2(t) = 1 - A_1(t)/A_2(t)$ and $A_2(t)/A_1(t) = \frac{\varphi''(r)}{\varphi'(r)(\varphi'(r)-\xi)} + 1$, with $r = \log t$ and A_i and B_i as defined above.

$\xi > 0$: In this case, $A_2(t)/A_1(t) \to \rho/\xi + 1$ as $t \to \infty$ and, consequently, the rates under linear and power norming, $A_2(t)$ and $B_2(t)$ respectively, are both of the same order $O(A_1(t))$, $t \to \infty$, as long as $\rho < 0$. For $\rho = 0$ however, $B_2(t) = o(A_2(t))$, $t \to \infty$, i.e the convergence rate is (asymptotically) faster when using power norming instead of linear norming.

$\xi = 0$: Assumption (A3) implies $\varphi''/\varphi' \to \rho \leq 0$, so that under (A1) a non-degenerate second-order relation (14) is only possible for $\rho = 0$. In this case, $\varphi''/(\varphi')^2 \to 0$ (by l'Hôpital's rule) and thus the convergence rate $B_2(.)$ for power normalized quantiles is (asymptotically) improved over the rate for linearly normalized quantiles.

Moreover, note that for $\rho = 0$ (and $\xi > 0$) the rate under a non-constant power normalization $b_2(t) = \varphi'(\log t)$ (*local slope*) is also faster than with a constant normalization $b(t) \equiv \xi = \varphi'(\infty)$ (*ultimate slope*). This in particular motivates the consideration of so-called *penultimate approximations* and their application to QRM.

Remark 3.2. (*penultimate approximation.*) From an asymptotic point of view, in the case $\xi = 0$ it matters whether we consider $U(tx)/U(t)$ together with a power normalization $1/b_2(t) = 1/\varphi'(\log t)$ and convergence to a constant limit x as in relation (12), or convergence of $U(tx)/U(t)$ to the non-constant, threshold-dependent limit $x^{b_2(t)}$. While in the former case the convergence rate is φ''/φ' (see Propsition 3.5), the rate in the latter case is φ''; see Degen and Embrechts [7], Theorem 3.2 (which also holds for $\xi = 0$ as one easily verifies). Clearly this difference is merely of theoretical interest, since, with QRM applications in mind, both procedures give rise to the same *penultimate* approximation $U(tx) \approx x^{b_2(t)} U(t)$ for $x > 1$ and t large. For an introduction to the concept of local slopes and penultimate approximations as well as references for the latter, see Degen and Embrechs [7].

Case III: $b_3(t) = \log U(t) - 1/t \int_{t_0}^{t} \log U(s) ds$

Assumption (A4) may be rewritten as $B_2(t)/B_3(t) - 1 \to \tau \leq 0$ for $t \to \infty$; see proof of Proposition 3.7. For $\tau \neq -1$, this implies that Condition (15) holds, i.e. $B_3(t)/A_3(t) \to \lambda$ for some $\lambda \in \mathbb{R}$. For the special case $\tau = -1$ which may lead to $\lambda = \infty$ in (15), we refer to Vanroelen [22] for further reading. As a consequence, for $\tau \neq -1$ and due to the second-order asymptotics in Case II, $B_3(.)$ and $A_3(.)$ are of the same order if and only if the second-order parameter satisfies $\rho < 0$. Similar to Case II, for $\rho = 0$, the rate under power norming may be improved over the rate under linear norming, i.e $B_3(t) = o\,(A_3(t))$, $t \to \infty$. Also, the rate under a non-constant normalization $b_3(.)$ ist faster than with constant norming $b(t) \equiv \xi$ (for $\xi > 0$). With applications in mind, this motivates the consideration of a *penultimate approximation* given by $U(tx) \approx x^{b_3(t)}U(t)$ for $x > 1$ and t large.

From the above we may draw the following conclusions about an EVT-based approximation of high quantiles under linear and power normalization. There are indeed situations where the convergence rate for quantiles with non-constant power normalization improves both the rates for linear normalization as well as constant power normalization. This in particular necessitates the second-order parameter ρ associated with the underlying model to satisfy $\rho = 0$ which pertains to many loss models important for QRM-practice. Hence we may hope for an improvement of the accuracy of classical EVT-based high-quantile estimators by the use of penultimate approximations.

In cases where the rate may be improved under power norming, the improvement does not seem to be spectacular at first glance as the improved rate is again slowly varying and hence may still be arbitrarily slow. However, the above second-order statements are about the *asymptotic behavior* of quantiles $F^{\leftarrow}(\alpha)$ only, i.e. as the confidence level α tends to 1. Of greater interest for QRM practice is the *local behavior* as one usually considers a fixed level $\alpha = 99.9\%$, say. Having said that, slow convergence (i.e. $\rho = 0$) does not at all need to be an impediment. In ranges relevant for practice the improvement in the estimation of high quantiles may well be considerable; see Degen and Embrechts [7], Figure 1.

4. Implications for quantitative risk management

We discuss the relevance of power norming, or more precisely of the corresponding penultimate approximations, for practical applications to QRM. In particular we study the EVT-based estimation of high quantiles together with possible fallacies it may bring with it. We hope that for the EVT-community, our discussion will lead to further relevant research—especially for the important case $\rho = 0$.

Recall the Basel II regulatory guidelines for CR and OR according to which risk capital has to be calculated using VaR (i.e. quantiles) at the high level of 99.9%. While this is standard for CR, where a variety of by now well-established models has been tried and tested, it is not so for OR. The latter has only been incorporated in the Basel II framework relatively recently, so that the resulting lack of historical data makes the estimation of a 99.9% quantile a daunting and inherently difficult task. Estimation methods include for instance the use of empirical quantiles as well as the fitting of some specific parametric loss models. For the latter method one is usually left with a reasonably good fit in the body but not in the tails of the data.

Due to the nature of the problem, the use of EVT has emerged naturally; see for instance Moscadelli [16], where the application of the popular Peaks Over Threshold (POT) method to OR is discussed. Having an extreme quantile level in mind, level scaling inherent to standard EVT (estimate at lower levels, 99% say, and then scale up to the desired higher levels such as 99.9%) provides a potential alternative. In either case however, accurate estimation of the tail index ξ is challenging, so that, in the end some constructive scepticism concerning the wiseness to base risk capital on high-level quantiles of some (profit and) loss df, even when using standard EVT methods, is still called for; see for instance Daníelsson et al. [6] and Nešlehová et al. [17].

The second-order results on power norming suggest that moving away from the tail index ξ—the indicator of the *ultimate* heavy-tailedness of the loss model—and focusing instead on the *local* tail index $b(t) = \varphi'(\log t)$, or on its pseudo equivalent $b(t) = \widetilde{\varphi}'(\log t)$, might prove useful at this point. In particular it motivates the consideration and comparison of estimation methods for high quantiles based on

i) standard EVT, and

ii) "advanced" EVT.

As for i), we incorporate two methods belonging to the standard EVT toolkit. Recall from the asymptotics for quantiles under linear norming (see relation (11)) that we may consider

$U(tx) \approx U(t) + a(t)\frac{x^\xi - 1}{\xi}$ and, due to regular variation of U, also $U(tx) \approx x^\xi U(t)$ for $x > 1$ and large values of t. This suggests the following scaling properties of high-quantile estimators. For some quantile levels $\widetilde{\alpha}, \alpha \in (0,1)$ with $\widetilde{\alpha} < \alpha$,

$$\widehat{\mathrm{VaR}_\alpha} = \widehat{\mathrm{VaR}_{\widetilde{\alpha}}} + \widehat{a(t)}\,\frac{x^{\widehat{\xi}} - 1}{\widehat{\xi}}, \tag{16}$$

and similarly

$$\widehat{\mathrm{VaR}_\alpha} = x^{\widehat{\xi}}\,\widehat{\mathrm{VaR}_{\widetilde{\alpha}}}, \tag{17}$$

with $x = (1-\widetilde{\alpha})/(1-\alpha) > 1$ and some estimates of ξ, $a(t)$ and $\mathrm{VaR}_{\widetilde{\alpha}}$ at the lower level $\widetilde{\alpha}$.

Relation (16) is better known as the *POT-estimator* of VaR_α. Indeed, setting $u = \widehat{\mathrm{VaR}_{\widetilde{\alpha}}}$, and using Proposition 2.1, we arrive at a natural estimator

$$\widehat{\mathrm{VaR}_\alpha} = u + \widehat{f(u)}\,\frac{\left(\frac{N_u}{n(1-\alpha)}\right)^{\widehat{\xi}} - 1}{\widehat{\xi}}, \tag{18}$$

for some estimates $\widehat{\xi}$ and $\widehat{f(u)}$ of ξ and of $f(u)$. Here $\frac{N_u}{n}$ is an estimate of $\overline{F}(u)$, where N_u denotes the number of exceedances over the threshold u (set by the user) of a total number of n data points; see for instance Embrechts et al. [8], Chapter 6.5.

In the simulation study below, (18) and (17) are referred to as the *Standard EVT I* and *II* methods, respectively. The tail index ξ and (threshold-dependent) scale parameter $f(u)$ are estimated using the POT-MLE method with an ad-hoc threshold choice of 10% of the upper order statistics. Compared to the POT-MLE, the performance of other implemented tail index estimators such as the Hill, the method of moments, and the exponential regression model (see for instance Beirlant et al. [3]) did not show significant differences.

Method ii) makes use of *penultimate* approximations. Based on relation (12), with a non-constant power normalization $b(.)$, we suggest the following scaling procedure for high-quantile estimators. For quantile levels $\widetilde{\alpha}, \alpha \in (0,1)$ with $\widetilde{\alpha} < \alpha$,

$$\widehat{\mathrm{VaR}_\alpha} = x^{\widehat{b(t)}}\,\widehat{\mathrm{VaR}_{\widetilde{\alpha}}}, \tag{19}$$

with $t = 1/(1-\widetilde{\alpha})$, $x = (1-\widetilde{\alpha})/(1-\alpha) > 1$ and some estimates of $b(t)$ and $\mathrm{VaR}_{\widetilde{\alpha}}$. For the simulation study, we incorporate the two choices $b(t) = \varphi'(\log t)$, the local slope, as well as

$b(t) = \widetilde{\varphi}'(\log t)$, the local "pseudo" slope, and will refer to these methods as the *Advanced EVT I* and *II* methods, respectively.

The advanced EVT methods, are included in the simulation study in order to outline the potential of penultimate approximations for practical applications. For the aim of this paper, we do not elaborate on the respective estimation procedures for φ' and $\widetilde{\varphi}'$. In both cases, the estimates are based on a prior local regression procedure for the log-data. This is done with the 'locfit' function (with a tricube weight function and smoothing parameter of $3/4$) provided in S-Plus (see Loader [14], Chapter 3 and Section 6.1). The integral appearing in $\widetilde{\varphi}'$ is approximated by a composite trapezoidal rule. Finally, the (lower) quantile $\mathrm{VaR}_{\widetilde{\alpha}}$ for (17) and (19) is estimated by the empirical quantile.

Remark 4.1. (*Local tail index.*)

The two scaling procedures (17) and (19) use the idea of a linear extrapolation of the log-log plot φ of U, but with slopes φ' at different quantile levels. While the penultimate approximation (19) requires the estimation of the local tail index $\varphi'(\log t)$ (or of $\widetilde{\varphi}'(\log t)$) at a specified levels t, the ultimate approximation (17)—in theory—makes use of estimates of the ultimate tail index $\varphi'(\infty) = \xi$.

In practice, given a sample of size a thousand, say, one will use a number of largest order statistics (above a certain threshold t_0) to estimate ξ in (17). It is clear that this yields an estimate of $\varphi'(\log u)$ at some (unknown) level $u > t_0$ rather than of $\xi = \varphi'(\infty)$. One of the differences between (17) and (19) thus is, that in the former case the level u is random (u depends on the underlying data), while the latter case uses estimates of the slope $\varphi'(\log t)$ at predefined levels $t = 1/(1 - \widetilde{\alpha})$, set by the user.

4.1. Simulation study

The simulation study is based on sample data from six frequently used OR loss models, such as the loggamma, the lognormal, the g-and-h, the Pareto, the Burr and the generalized Beta distribution of the second kind (GB2). For convenience we recall the definition of a g-and-h rv

X which is obtained from a standard normal rv Z through

$$X = a + b\frac{e^{gZ} - 1}{g} e^{hZ^2/2},$$

with parameters $a, g, h \in \mathbb{R}$ and $b \neq 0$. Note that in the case $h = 0$ one obtains a (shifted) lognormal rv. For the Pareto df we use the parameterization $\overline{F}(x) = (x/x_0)^{-1/\xi}$, for $x > x_0 > 0$ and some $\xi > 0$. The GB2 is parameterized as in Kleiber and Kotz [13], p. 184, while the remaining three loss models are as in Embrechts et al. [8], p. 35.

For Table 1 we simulate 200 samples of 1000 observations from each of the six loss models. For each of the 200 data sets we compare bias and the standardized root mean square error (SRMSE) of the four above-mentioned EVT-based estimation methods for VaR at level 99.9%. Several simulations with different choices of (for risk management practice relevant) parameter values were performed, all of them showing a similar pattern concerning the performance of the different estimation methods; see Table 1.

TABLE 1: Bias and SRMSE (in %) of four EVT-based estimators for VaR at the 99.9% level based on based on 200 datasets of 1000 observations of six different loss models.

	Bias	SRMSE	Bias	SRMSE	Bias	SRMSE
Loss model	Loggamma ($\alpha = 1.75, \beta = 2$)		Lognormal ($\mu = 3.5, \sigma = 1.25$)		g-and-h ($a = b = 3$, $g = 0.8, h = 0.4$)	
Std. EVT I (POT)	8.41	52.88	5.20	32.93	9.65	57.63
Std. EVT II ($\widetilde{\alpha} = 0.99$)	5.26	56.53	-8.88	39.24	4.97	62.62
Adv. EVT I ($\widetilde{\alpha} = 0.99$)	5.69	35.51	14.34	35.23	7.77	44.80
Adv. EVT II ($\widetilde{\alpha} = 0.99$)	7.60	36.84	42.44	53.21	9.53	44.36
	Pareto ($x_0 = 1.2, \xi = 0.75$)		Burr ($\alpha = 1, \kappa = 2, \tau = 1.5$)		GB2 ($a = b = 2$, $p = 1.5, q = 0.75$)	
Std. EVT I (POT)	13.73	62.73	7.79	54.12	1.20	45.80
Std. EVT II ($\widetilde{\alpha} = 0.99$)	13.99	72.48	6.10	62.20	0.21	51.65
Adv. EVT I ($\widetilde{\alpha} = 0.99$)	-9.53	28.29	1.98	41.34	-5.10	29.94
Adv. EVT II ($\widetilde{\alpha} = 0.99$)	2.66	41.95	3.60	39.80	-1.69	32.35

Remark 4.2. Despite its inconsistency with the well-known stylized facts of OR data (power-tail, i.e. $\xi > 0$), the lognormal distribution (semi heavy-tailed, i.e. $\xi = 0$) is widely used in OR practice as a loss severity model. We include it in our simulation study primarily to question its omnipresence by highlighting some of the problems its use may bring with it.

As mentioned above, estimation at very high quantile levels by means of fitting a parametric loss model may be hard to justify. For illustrative purposes we nevertheless perform a simulation for the six resulting parametric high-quantile estimators, based on the same data sample. An excerpt of these (expectedly) disappointing results is given in Table 2. Here, the model parameters are estimated using MLE, except for the g-and-h distribution, for which there is no agreed standard estimation method so far. For that case we adapt a method suggested by Tukey [21] based on $\log_2 n$ so-called letter values, where n is the sample size.

Table 2: Bias and SRMSE (in %) of parametric estimators for VaR at the 99.9% level based on based on 200 datasets of 1000 observations of three different loss models.

Loss model	Bias	SRMSE	Bias	SRMSE	Bias	SRMSE
	Lognormal ($\mu = 3.5, \sigma = 1.25$)		Burr ($\alpha = 1, \kappa = 2, \tau = 1.5$)		GB2 ($a = b = 2$, $p = 1.5, q = 0.75$)	
Loggamma	703.51	735.81	188.78	200.70	72.59	81.21
Lognormal	0.50	9.38	-57.86	58.08	-74.88	74.92
g-and-h	-4.27	15.57	-45.33	47.59	-45.46	47.03
Pareto	1.04e+13	8.51e+13	7.87e+19	1.029e+21	2.57e+10	2.33e+11
Burr	-89.77	89.81	1.69	26.73	20.12	34.35
GB2	91.42	300.91	1.26	32.09	-2.00	25.36

A comparison of the results in the Tables 1 and 2 clearly shows that the estimation of high quantiles based on fitting parametric models may indeed be problematic. The model uncertainty involved may be considerable (large fluctuation of the estimation errors). Moreover, from a QRM regulatory point of view, a large negative bias (i.e. underestimation of risk capital) is to be avoided. Not surprisingly, the lognormal parametric model underestimates risk capital charges considerably. While intolerable from a sound regulatory perspective this at the same time may explain the "attractiveness" of its use for a financial institution.

On the other hand, given the high level of 99.9%, the performance of all four EVT-based methods is promising; see Table 1. A comparison within the EVT-based methods does not yield a clear ranking. However, the advanced EVT methods seem to work at least as well as the standard EVT methods, in particular exhibiting smaller SRMSE. This finding is not by accident. Recall that the estimation of φ' and $\widetilde{\varphi}'$ in the advanced EVT I and II methods is based on a local regression procedure (i.e. smoothing) of the log-data. As a consequence, the estimates are more robust, which leads to smaller SRMSE-values. For smaller sample sizes we expect this behavior to become even more pronounced.

To confirm the above findings on EVT-based high-quantile estimators, we perform a second, similar study and estimate quantiles at the even more extreme level of 99.97%, relevant for the calculation of so-called economic capital; see for instance Crouhy et al. [5], Chapter 15. Owing to Remark 4.2 we leave out the lognormal data sample. We again simulate 200 samples of 1000, 500 and 250 observations of very heavy-tailed data in Table 3.

From Table 3 we may draw the following conclusions. Most importantly, the potential of an advanced EVT approach to estimate extreme quantiles in the presence of very heavy tails and small sample sizes is clearly revealed. The performance of the advanced EVT I and II methods is by far superior compared to that of the two standard EVT approaches. This confirms that the idea of using penultimate approximations instead of ultimate approximations may indeed be promising in certain situations relevant for practice (and not only from a second-order asymptotic viewpoint). While the estimation errors of the two advanced EVT methods remain comparably moderate even for small sample sizes, standard EVT-based methods hit the wall. The estimation errors explode for decreasing sample sizes so that the usefulness of these methods seems questionable in such situations. From a QRM perspective this means that relying on high-quantile estimates based on these conventional methods may be questionable.

Table 3: Bias and SRMSE (in %) of four EVT-based estimators for VaR at the 99.97% level based on 200 datasets of 1000, 500 and 250 observations.

	$n = 1000,\ \widetilde{\alpha} = 0.99$		$n = 500,\ \widetilde{\alpha} = 0.98$		$n = 250,\ \widetilde{\alpha} = 0.96$	
	Bias	SRMSE	Bias	SRMSE	Bias	SRMSE
Loggamma ($\alpha = 1.25, \beta = 1.25$)						
Std. EVT I (POT)	39.47	159.44	81.57	265.64	839.68	8934.55
Std. EVT II	38.19	160.53	82.15	277.51	1150.21	11944.19
Adv. EVT I	-2.99	46.88	-3.93	54.19	-7.73	65.91
Adv. EVT II	7.49	68.89	1.94	65.52	-14.11	80.61
g-and-h ($a = b = 1.5, g = 0.8, h = 0.6$)						
Std. EVT I (POT)	43.06	149.69	80.63	251.15	257.08	963.06
Std. EVT II	39.94	163.40	84.14	278.85	362.78	1426.99
Adv. EVT I	7.76	60.52	16.76	75.44	40.31	130.65
Adv. EVT II	17.52	83.57	18.38	92.22	8.62	121.65
Pareto ($x_0 = 1, \xi = 0.85$)						
Std. EVT I (POT)	33.31	120.47	105.22	317.70	176.93	1112.75
Std. EVT II	35.14	135.80	118.95	354.66	265.77	1734.51
Adv. EVT I	-16.29	35.67	-29.95	43.54	-31.36	53.36
Adv. EVT II	5.46	63.49	-8.24	71.91	-22.20	65.45
Burr ($\alpha = 1, \kappa = 1.5, \tau = 1.25$)						
Std. EVT I (POT)	29.94	159.70	68.72	263.39	244.88	1474.04
Std. EVT II	27.77	166.73	68.98	285.69	287.82	1566.36
Adv. EVT I	5.29	69.86	24.87	88.72	81.04	207.97
Adv. EVT II	9.26	75.01	16.09	79.27	19.82	99.54
GB2 ($a = 1, b = 2, p = 1.5, q = 1.25$)						
Std. EVT I (POT)	12.93	88.16	104.19	589.04	143.92	613.16
Std. EVT II	11.63	93.63	108.70	661.79	207.61	970.47
Adv. EVT I	6.58	58.63	29.20	97.35	95.53	245.15
Adv. EVT II	12.96	59.20	24.79	81.35	49.89	144.99

5. Conclusion

In this paper we consider EVT-based high-quantile estimators and discuss scaling properties and their influence on the estimation accuracy at very high quantile levels. The scarcity of data together with the heavy-tailedness present in the data (especially for OR), turns high-quantile estimation into an inherently difficult statistical task. The nature of the problem calls for EVT in some or other form. The application of methods from the standard EVT toolkit in such applied situations is however not without problems. Our main results are as follows.

First, from a methodological perspective, it is de Haan's framework of Π-variation that is most useful for our purposes, as it allows for a unified treatment of the for QRM important cases $\xi > 0$ and $\xi = 0$. Inherent to Π-variation is the notion of power norming (as opposed to the standardly used linear norming) of quantiles and high-risk scenarios. The use of different normalizations leads to different second-order asymptotics. It turns out that, in certain cases relevant for practice, judicious choices of a (non-constant) power normalization—instead of a linear or a constant power normalization—may improve the rate of convergence in the respective limit results.

Second, the theory of second-order extended regular variation provides a methodological basis for the derivation of new high-quantile estimators. The application of different normalizations in the respective second-order relations translates into different scaling properties of the resulting high-quantile estimators. Our findings motivate the derivation of new estimation procedures for high quantiles by means of penultimate approximations. In particular we propose two "advanced" EVT methods which are based on the estimation of the local (pseudo) slope φ' (and $\widetilde{\varphi}'$) of the log-log plot φ of the underlying loss model $U(t) = e^{\varphi(\log t)}$. The methods proposed are intended to complement, rather than to replace, methods from the standard EVT toolkit. Their applications may be useful in situations in which the reliability of standard methods seems questionable.

Third, by means of a simulation study we show that, in the presence of heavy tails together with data scarcity, reliable estimation at very high quantile levels, such as the 99.9% or 99.97%, is a tough call. While our study highlights the limitations of standard EVT approaches in such cases, at the same time it reveals the potential of more advanced EVT methods.

Further statistical research on advanced EVT approaches to estimate high quantiles, together with a more in-depth study of their benefits as well as limitations for practical applications would be desirable.

6. Acknowledgements

The authors would like to thank Guus Balkema for useful discussions related to the paper.

References

[1] Balkema, G. and Embrechts, P. (2007). *High Risk Scenarios and Extremes - A geometric approach.* EMS Publishing House, Zürich.

[2] Barakat, H. M., Nigm, E. M. and El-Adll, M. E. (2008). Comparison between the rates of convergence of extremes under linear and under power normalization. Statistical Papers, Springer. DOI: 10.1007/s00362-008-0128-1.

[3] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004). *Statistics of Extremes.* Wiley, Chichester.

[4] Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987). *Regular Variation.* Cambridge University Press, Cambridge.

[5] Crouhy, M., Galai, D. and Mark, R. (2006). *The Essentials of Risk Management.* McGraw-Hill, New York.

[6] Daníelsson, J., Embrechts, P., Goodhart, C., Keating, C., Muennich, F., Renault, O. and Song Shin, H. (2001). An academic response to Basel II. Financial Markets Group, London School of Economics.

[7] Degen, M. and Embrechts, P. (2008). EVT-based estimation of risk capital and convergence of high quantiles. *Advances in Applied Probability* **40,** 696–715.

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

[9] Fisher, R. A. and Tippett, L. H. T. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Camb. Phil. Soc.* **24,** 180–190.

[10] Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics* **44,** 423–453.

[11] de Haan, L. (1970). On regular variation and its applications to the weak convergence of sample extremes. CWI Tract 32, Amsterdam.

[12] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory - An Introduction.* Springer, New York.

[13] Kleiber, C. and Kotz, S. (2003). *Statistical Size Distributions in Economics and Actuarial Sciences.* Wiley, Hoboken.

[14] Loader, C. (1999). *Local Regression and Likelihood.* Springer, New York.

[15] Mohan, N. R. and Ravi, S. (1991). Max domains of attraction of univariate and multivariate p-max stable laws. *Theory Probab. Appl.* **37,** 632–643.

[16] Moscadelli, M. (2004). The modelling of operational risk: experiences with the analysis of the data collected by the Basel Committee. Bank of Italy, Working Paper No 517.

[17] Nešlehová, J., Embrechts, P. and Chavez-Demoulin, V. (2006). Infinite mean models and the LDA for operational risk. *Journal of Operational Risk* **1,** 3–25.

[18] Pantcheva, E. (1985). Limit theorems for extreme order statistics under nonlinear normalization. *Lecture Notes in Mathematics, No. 1155,* 284–309. Springer, Berlin.

[19] Resnick, S. I. (1987). *Extreme Values, Regular Variation and Point Processes.* Springer, New York.

[20] Basel Committee on Banking Supervision (2008). Guidelines for Computing Capital for Incremental Risk in the Trading Book. Basel: Bank for International Settlements.

[21] Tukey, J. W. (1977). *Exploratory Data Analysis.* Addison-Wesley, Reading.

[22] Vanroelen, G. (2003). The effect of transformations on second-order regular variation. Ph.D. Thesis, Katholieke Universiteit Leuven.