Applied Probability Trust (January 27, 2011)

SCALING OF HIGH-QUANTILE ESTIMATORS

MATTHIAS DEGEN,* **

PAUL EMBRECHTS,** *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. [8]. Also the weaknesses of Value-at-

** Postal address: Department of Mathematics, ETH Zurich, Raemistrasse 101, CH-8092 Zurich, Switzerland. Email address: embrechts@math.ethz.ch

* Current address: School of Operations Research and Information Engineering, Cornell University, Rhodes Hall 279, Ithaca, NY 14853, USA. Email address: md555@cornell.edu

Risk (VaR), the risk measure required by the Basel framework, were discussed over and over again; see for instance Nešlehová et al. [20] 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/III 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 [3]. 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. In a second part, concrete applications of the methodology developed in the first part are discussed.

The paper is organized as follows. In Section 2 we introduce some basic concepts from EVT. In Section 3 we discuss the concept of normalized high-risk scenarios and, in Section 4, compare the effects of linear versus power norming for high-risk scenarios and quantiles using the framework of first- and second-order extended regular variation. Based on the findings from these asymptotic results, we propose the use of so-called *penultimate approximations* to estimate extreme quantiles. In Section 5 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. [10] or in de Haan and Ferreira [15]. 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; For properties of $F^{\leftarrow}$, see for instance Embrechts and Hofert [9]. 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 [15], 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

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

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 [15], 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 [12]. The novelty of Proposition 2.1 (originally due to de Haan [14]) 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 [15], 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. First-order asymptotics 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 [21] and Mohan and Ravi [18] 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 $\left(X^t/t\right)^{1/g(t)}$ 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.

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 3.1. (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 $x > 0$

$$\frac{\overline{F}\left(tx^{g_1(t)}\right)}{\overline{F}(t)} \to \overline{K}_1(x), \quad \frac{\overline{F}\left(tx^{g_2(t)}\right)}{\overline{F}(t)} \to \overline{K}_2(x), \quad t \to x_F, \tag{6}$$

then

$$\lim_{t \to x_F} \frac{g_2(t)}{g_1(t)} = p > 0 \tag{7}$$

and

$$K_2(x) = K_1(x^p). \tag{8}$$

ii) If (7) holds, then either of the two relations in (6) implies the other and (8) holds.

Proof. ii) Assume that (7) 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 (6) 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 (see Resnick [22], Proposition 0.1), we have that

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

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 the result below we exploit the link between the two concepts of linear and power norming for high-risk scenarios. It connects the respective domains of attraction D_l^{POT} and D_p^{POT} and may be seen as a consequence of the classical Convergence to Types Theorem (see Resnick [22], Proposition 0.2) so that we refrain from giving a proof here.

Proposition 3.2. *For $X > 0$ with df F_X and for $\xi \in \mathbb{R}$ the following holds:*

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

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

As we subsequently prefer to work within a quantile setting, a reformulation of Proposition 3.2 in terms of quantile functions is useful.

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

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

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

Remark 3.1. The respective converse implications in ii) of Proposition 3.2 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 $F_X \in D_p^{POT}$.

4. Second-order asymptotics of normalized quantiles

The results below are expressed in terms of 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 $\xi \geq 0$, of most interest for applications in insurance and finance. Similar results for the case $\xi < 0$ may be worked out.

Assuming $U \in ERV_\xi(a)$ for some $\xi \geq 0$, i.e. for $x > 0$

$$\frac{U(tx) - U(t)}{a(t)} \to D_\xi(x) := \frac{x^\xi - 1}{\xi}, \quad t \to \infty, \tag{9}$$

for some measurable function $a(.) > 0$, Corollary 3.1 implies $\log U \in \Pi(b)$ and hence

$$\left(\frac{U(tx)}{U(t)}\right)^{1/b(t)} \to x, \quad t \to \infty, \tag{10}$$

where $b(t) = a(t)/U(t) > 0$ and such that $b(t) \to \xi$. As a consequence, the (high) quantile $U(tx)$ may for large values of t either be approximated by

$$U(tx) \approx U(t) + a(t)D_\xi(x) \tag{11}$$

or by

$$U(tx) \approx x^{b(t)}U(t). \tag{12}$$

While the former approximation is well-studied (see for instance de Haan and Ferreira [15], Section 3), the latter is less known and hence of main interest in the sequel. The two approximations (11) and (12) will in general yield different results (unless $b(t) \equiv \xi$ for some $\xi > 0$ in which case they coincide). In order to exploit the potential of Approximation (12) we compare its performance with the standardly used Approximation (11) by means of comparing the respective relative approximation errors in an asymptotic framework, followed by a simulation study in Section 5.

Proposition 4.1. *Suppose there exist functions b, with $\lim_{t\to\infty} b(t) = \xi$ for some $\xi \geq 0$, and B, ultimately monotone and with $\lim_{t\to\infty} B(t) = 0$ such that for some $\rho \leq 0$ and for $x > 0$,*

$$\lim_{t\to\infty} \frac{\frac{x^{b(t)}U(t)}{U(tx)} - 1}{B(t)} = -T_\rho(x), \tag{13}$$

where

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

In the case that $\xi = \rho = 0$, we further assume that $\lim\limits_{t\to\infty} B(t)/(b(t))^2 = c \in \mathbb{R}$. Then we have that for $x > 0$

$$\lim_{t\to\infty} \frac{\frac{U(t)+a(t)D_\xi(x)}{U(tx)} - 1}{A(t)} = -S_{\xi,\rho}(x), \tag{14}$$

where $a(t) = b(t)U(t)$ and

$$A(t) = \begin{cases} b(t) - \xi, & \rho = 0 < \xi, \\ B(t), & \rho < 0 \text{ or } (\xi = \rho = 0 \text{ and } c \neq 0), \\ (b(t))^2, & \xi = \rho = 0, c = 0. \end{cases}$$

and with

$$S_{\xi,\rho}(x) = \begin{cases} \frac{1}{\rho}\left(\log x - x^{-\xi}D_\xi(x)\right) + T_\rho(x), & \rho < 0, \\ \log x - x^{-\xi}D_\xi(x), & \rho = 0 < \xi, \\ \left(1 + \frac{1}{c}\right)T_0(x), & \xi = \rho = 0, c \neq 0, \\ T_0(x), & \xi = \rho = 0, c = 0. \end{cases}$$

Proof. First note that (13) can for $x > 0$ be rewritten as

$$\lim_{t\to\infty} \frac{-\left(\frac{x^{b(t)}U(t)}{U(tx)} - 1\right)}{B(t)} = \lim_{t\to\infty} \frac{\log U(tx) - \log U(t) - b(t)\log x}{B(t)} = T_\rho(x). \tag{15}$$

Since $\lim_{t\to\infty} B(t)/b(t) = 0$ we thus have that $\log U \in 2ERV_{0,\rho}(b, B/b)$; see for instance de Haan and Ferreira [15], Appendix B.3 for an introduction to 2ERV. The assumed form of the limit $T_\rho(x)$ implies that $b \in ERV_\rho(B)$ and hence also $(b(t) - \xi) \in ERV_\rho(B)$. Moreover, we have that

$$\lim_{t\to\infty} \frac{B(t)}{b(t) - \xi} = \rho, \quad (\rho \leq 0 \leq \xi); \tag{16}$$

see de Haan and Ferreira [15], Theorem B.2.2 and Corollary B.2.13.

Now, let $\xi > 0$ and observe that as $t \to \infty$,

$$\begin{aligned}
-\left(\frac{U(t)+a(t)D_\xi(x)}{U(tx)} - 1\right) &\sim \frac{U(tx)}{U(t)(1+b(t)D_\xi(x))} - 1 \sim x^{-\xi}\left(\frac{U(tx)}{U(t)} - 1 - b(t)D_\xi(x)\right) \\
&= \frac{U(tx)}{x^\xi U(t)} - 1 - (b(t)-\xi)x^{-\xi}D_\xi(x) \\
&= (\log U(tx) - \log U(t) - \xi \log x)\,(1+o(1)) - (b(t)-\xi)x^{-\xi}D_\xi(x) \\
&= b(t)\log x + T_\rho(x)B(t) - \xi\log x - (b(t)-\xi)x^{-\xi}D_\xi(x) \\
&\quad + o(B(t)) + o(b(t)-\xi) \\
&= \left(\log x - x^{-\xi}D_\xi(x)\right)(b(t)-\xi) + T_\rho(x)B(t) + o(B(t)) + o(b(t)-\xi),
\end{aligned}$$

where we used a Taylor expansion for exp(.) and (15). Therefore we have

$$\begin{aligned}
\frac{\frac{U(t)(1+b(t)D_\xi(x))}{U(tx)} - 1}{A(t)} &\sim -\left(\log x - x^{-\xi}D_\xi(x)\right)\frac{b(t)-\xi}{A(t)} - T_\rho(x)\frac{B(t)}{A(t)} \\
&\quad + \left(o(B(t)) + o(b(t)-\xi)\right)\frac{1}{A(t)},
\end{aligned}$$

so that the result follows in view of (16).

Now let $\xi = 0$ and recall that $\lim_{t\to\infty} B(t)/b(t) = 0$. On the other hand, $b \in ERV_\rho(B)$ implies $B(t)/b(t) \to \rho$ and therefore the case $\xi = 0$ necessitates $\rho = 0$. Then, as $t \to \infty$,

$$\begin{aligned}
-\left(\frac{U(t)(1+b(t)D_\xi(x))}{U(tx)} - 1\right) &\sim \log U(tx) - \log U(t) - \log\left(1 + b(t)\log x\right) \\
&= b(t)\log x + T_0(x)B(t) + o(B(t)) \\
&\quad - \left(b(t)\log x - \frac{1}{2}(\log x)^2(b(t))^2 + o((b(t))^2)\right)
\end{aligned}$$

and hence

$$\frac{\left(\frac{U(t)(1+b(t)D_\xi(x))}{U(tx)} - 1\right)}{A(t)} \sim -T_0(x)\frac{B(t)}{A(t)} - \frac{1}{2}(\log x)^2\frac{(b(t))^2}{A(t)} + \frac{o(B(t)) + o((b(t))^2)}{A(t)},$$

which finishes the proof.

Remarks 4.1.

i) From Proposition 4.1 and its proof we may conclude that the (less known) Approximation (12) performs asymptotically at least as good as Approximation (11). Indeed, in the case $\rho < 0$ the approximation errors tend to zero at the same rate $B(t)$ (except for the special case $-\rho = \xi > 0$ for which $S_{\xi,-\xi}(x) \equiv 0$ or if $c = -1$). In the case $\rho = 0$ (and $c = 0$ if $\xi = 0$), the error rate in (12) tends to zero faster than in (11). This is of particular interest with focus on possible applications to quantitative risk management, where frequently used models (E. Balta, Office of the Comptroller of the Currency, personal communication) include for instance the lognormal ($\xi = \rho = 0$), the loggamma or the g-and-h (both $\rho = 0$).

ii) In cases where the relative approximation error of (12) vanishes faster than that of (11), the gain is not spectacular since these cases necessitate $\rho = 0$. The corresponding convergence rate $B(t)$ in (13) is slowly varying and thus may tend zo zero arbitrarily slow. Similar conclusions are found in de Haan and Gomes [13] in the context of penultimate approximations in the block-maxima setting.

iii) From a methodological viewpoint, Proposition 4.1 may be seen as a partial converse of Lemma B.3.16. of de Haan and Ferreira [15]. While these authors show how the assumption of $U \in 2ERV_{\xi,\rho}$ implies a second-order condition for $\log U$, we basically assume $\log U \in 2ERV_{0,\rho}$ and analyze the implications on the second-order behavior of U. Note that in their framework the case $\xi = \rho$ (e.g. lognormal) is not treated. Also, in the case $\rho = 0$ (e.g. loggamma, g-and-h) no non-degenerate second-order result for $\log U$ is obtained.

In summary, while Proposition 4.1 highlights the potential usefulness of Approximation (12), the findings are asymptotic and hence do not guarantee a good performance for finite samples. Therefore, numerical simulations are needed in order to evaluate the potential of (12) for practical applications. To do so, we must first identify candidates $b(.)$ satisfying (13). Below we consider two different choices of $b(.)$ and derive sufficient conditions for (13) to hold.

4.1. Sufficient Conditions

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 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 [15] 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''. With this notation introduced, we have the following result on sufficient conditions for Proposition 4.1 to hold.

Proposition 4.2. *Suppose $U(t) = e^{\varphi(\log t)}$ is three times differentiable and satisfies (A1).*

i) Let $b_1(t) = \varphi'(\log t)$ and assume that b_1' ultimately monotone and that $\lim_{t\to\infty} \varphi'''(t)/\varphi''(t) = \rho$, for some $\rho \leq 0$. Then (13) holds with $b(t) = b_1(t)$ and $B(t) = tb_1'(t) = \varphi''(\log t)$.

ii) Let $b_2(t) = \log U(t) - 1/t \int_{t_0}^{t} \log U(s)ds$, for some $t_0 > 0$, and assume that b_2' is ultimately monotone and that $\lim_{t\to\infty} \varphi'''(\log t)/\left(\varphi''(\log t) - tb_2'(t)\right) - 1 = \rho$, for some $\rho \leq 0$. If $\rho \neq -1$, then (13) holds with $b(t) = b_2(t)$, $B(t) = tb_2'(t)$ and with limit $T_\rho(x) + D_\rho(x)$.

Proof. For i), we may rewrite (13) for $x > 0$ and with $t \to \infty$ as

$$\begin{aligned}
\lim_{t\to\infty} \frac{-\left(\frac{x^{b_1(t)}U(t)}{U(tx)} - 1\right)}{B(t)} &= \lim_{t\to\infty} \frac{\log U(tx) - \log U(t) - b_1(t)\log x}{B(t)} \\
&= \int_1^x \frac{b_1(ts) - b_1(t)}{B(t)} \frac{1}{s} ds.
\end{aligned}$$

With b_1 as given, $\varphi'''/\varphi'' \to \rho$ is equivalent to $tb_1''(t)/b_1'(t) \to \rho - 1$ and together with ultimate monotonicity of b_1' ensures that $b_1 \in ERV_\rho(B)$ for some $\rho \leq 0$, such that we may choose $B(t) = tb_1'(t)$. While clear for the case $\rho < 0$, this follows from the Monotone Density Theorem for $\Pi-$Variation for the case $\rho = 0$; see Bingham et al. [5], Theorem 3.6.8. Finally, by the Uniform Convergence Theorem for ERV(see Bingham et al. [5], Theorem 3.1.7a), the convergence

$$\lim_{t\to\infty} \frac{b_1(ts) - b_1(t)}{tb_1'(t)} = \begin{cases} \frac{s^\rho - 1}{\rho}, & \rho < 0, \\ \log s, & \rho = 0. \end{cases}$$

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

The proof for the ii) is similar to i); the main steps are as follows. With b_2 as given, the assumptions guarantee that $b_2 \in ERV_\rho(c)$ with $c(t) = tb_2'(t)$. Furthermore, using partial integration one gets

$$\log U(x) = b_2(x) + \int_{t_0}^{x} \frac{b_2(t)}{t} dt.$$

Therefore, again by the Uniform Convergence Theorem for ERV, we obtain for $x > 0$ and as $t \to \infty$,

$$
\begin{aligned}
-\left(\frac{\frac{U(tx)}{x^{b_2(t)}U(t)} - 1}{tb_2'(t)}\right) &\sim \frac{\log U(tx) - \log U(t) - b_2(t)\log x}{tb_2'(t)} \\
&= \frac{b_2(tx) - b_2(t)}{tb_2'(t)} + \int_1^x \frac{b_2(ts) - b_2(t)}{tb_2'(t)}\frac{1}{s}ds \\
&\to \frac{x^\rho - 1}{\rho} + T_\rho(x),
\end{aligned}
$$

which finishes the proof.

The rationale behind the choice of normalization b_1 in Proposition 4.2 i) is rather intuitive once we note that $b_1(t) = tU'(t)/U(t) = \varphi'(\log t)$ is the slope of the log-log plot of $U(t) = e^{\varphi(\log t)}$. Under (A1), obviously $\varphi'(\log t) \to \varphi'(\infty) = \xi$, and hence we will refer to $b(t) = \varphi'(\log t)$ as the *local* or *penultimate tail index* of the log-log plot of U at points t (as opposed to the *ultimate* tail index $\varphi'(\infty) = \xi$). Further, we remark that the sufficient conditions given in Proposition 4.2 i) are—under suitable smoothness and monotonicity assumptions on U—close to also being necessary for (13).

In contrast, the choice of normalization b_2 in Proposition 4.2 ii) presents a special case. We include it in this asymptotic analysis merely to present an alternative way of choosing $b(.)$. This in turn will result in a different high-quantile estimator. Unlike b_1, the rationale behind b_2 does not seem to be very intuitive at first. It may be motivated by Karamata's Theorem, according to which $\varphi(\log t)$ 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 $t \to \infty$. Therefore, one may choose $b(t) = \widetilde{\varphi}'(\log t) = \varphi(\log t) - \frac{1}{t}\int_{t_0}^t \varphi(\log s)ds$ with $0 < t_0 < t$.

In view of the discussion above, we will refer to the approximation $U(tx) \approx x^{b(t)}U(t)$ for some $b(t) \to \xi \geq 0$ as the *penultimate* approximation (as opposed to the *ultimate* approximation $U(tx) \approx x^{\xi}U(t)$). The idea of penultimate approximations goes back to Fisher and Tippett [11]. Gomes and de Haan [13] for instance discuss penultimate approximations in the context of block maxima. The potential of penultimate approximations for practical applications seems to have received limited attention so far. Motivated by the asymptotic results above, below we analyze the potential of penultimate approximations for high-quantile estimation by means of a simulation study.

5. Implications for quantitative risk management

We discuss the relevance of power norming, or more precisely of the corresponding penultimate approximations as discussed in the previous section. 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/III 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%. Due to the nature of the problem, the use of EVT has emerged naturally; see Moscadelli [19] in the case of OR and Chavez-Demoulin and Embrechts [6] for CR. 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. [8] and Nešlehová et al. [20].

The asymptotic results discussed in Section 4 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 $b(t) = \widetilde{\varphi}'(\log t)$, might prove useful at this point. In particular it motivates the consideration and comparison of estimation methods for high quantiles based on what we would like to call i) standard EVT, and ii) advanced EVT (see below for more details on this nomenclature).

As for i), we incorporate two methods belonging to the standard EVT toolkit. Recall from the asymptotics for quantiles under linear norming (see relation (9)) 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 $\widetilde{\alpha} < \alpha$,

$$\widehat{\mathrm{VaR}_{\alpha}} = \widehat{\mathrm{VaR}_{\widetilde{\alpha}}} + \widehat{a(t)}\,\frac{x^{\widehat{\xi}}-1}{\widehat{\xi}}, \tag{17}$$

and similarly

$$\widehat{\mathrm{VaR}_{\alpha}} = x^{\widehat{\xi}}\,\widehat{\mathrm{VaR}_{\widetilde{\alpha}}}, \tag{18}$$

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 (17) 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

$$\widehat{\mathrm{VaR}_\alpha} = u + \widehat{f(u)}\,\frac{\left(\frac{N_u}{n(1-\alpha)}\right)^{\widehat{\xi}} - 1}{\widehat{\xi}}, \tag{19}$$

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. [10], Chapter 6.5.

In the simulation study below, (19) and (18) 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; extensive simulations (V. Chavez-Demoulin, personal communication) have shown that this is an overall good first threshold choice. 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. [4]) did not show significant differences.

The so-called advanced EVT approach ii) makes use of *penultimate* approximations. Based on relation (10), 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 $\widetilde{\alpha} < \alpha$,

$$\widehat{\mathrm{VaR}_\alpha} = x^{\widehat{b(t)}}\,\widehat{\mathrm{VaR}_{\widetilde{\alpha}}}, \tag{20}$$

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 $b(t) = \varphi'(\log t)$ as well as $b(t) = \widetilde{\varphi}'(\log t)$ 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 [17], 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 (18) and (20) is estimated by the empirical quantile.

Remark 5.1. (*Local tail index.*) The two scaling procedures (18) and (20) 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 (20) 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 (18)—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 (18). 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 (18) and (20) 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.

5.1. 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 [16], p. 184, while the remaining three loss models are as in Embrechts et al. [10], p. 35.

For Table 1 we simulate 200 samples of 1000 observations from each of the six loss models. For each of the four above-mentioned EVT-based estimation methods we then calculate estimates $(\widehat{q}^{(i)}_{0.999})_{1 \le i \le 200}$ of VaR at level 99.9% and compare the respective bias and the standardized root mean square error (SRMSE), which is defined as

$$\frac{1}{q_{0.999}} \sqrt{\frac{1}{200} \sum_{i=1}^{200} \left(\widehat{q}^{(i)}_{0.999} - q_{0.999}\right)^2}.$$

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 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 5.2. 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 [23] based on $\log_2 n$ so-called letter values, where n is the sample size.

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.

Table 2: Bias and SRMSE (in %) of parametric estimators for VaR at the 99.9% level 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

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. [7], Chapter 15. Owing to Remark 5.2 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 far superior to the two standard EVT approaches. This confirms that 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). The estimation errors of the two advanced EVT methods remain comparably moderate, even for small sample sizes. The estimation errors for standard EVT methods explode for small sample sizes. From a

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

QRM perspective this means that relying on high-quantile estimates based on these conventional methods may become questionable.

6. 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% remains a very difficult task. Regulators as well as practitioners ought to become more aware of this issue and consequently temper their aspiration of reaching very reliable capital estimation so far in the tail of loss distributions. While our study highlights the limitations of standard EVT approaches in such cases, given the above constraint, 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.

7. Acknowledgments

The authors would like to thank Guus Balkema for useful discussions related to the paper as well as the anonymous referees for their constructive remarks. Matthias Degen acknowledges financial support from the Swiss National Science Foundation (SNSF) and thanks ORIE at Cornell University for the hospitality. As Senior SFI Professor, Paul Embrechts thanks the Swiss Finance Institute (SFI) for financial support.

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

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

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

[6] Chavez-Demoulin, V. and Embrechts, P. (2010). An EVT primer for credit risk. In *Handbook of Credit Derivatives.* ed. A. Lipton and A. Rennie. Oxford University Press. To appear.

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

[8] 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.

[9] Embrechts, P. and Hofert, M. (2010). A note on generalized inverses. Preprint, ETH Zurich.

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

[11] 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.

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

[13] Gomes, M. I. and de Haan, L. (1999). Approximation by penultimate extreme value distributions. *Extremes* **2,** 71–85.

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

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

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

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

[18] 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.

[19] 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.

[20] 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.

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

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

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