Applied Probability Trust (February 13, 2008)

EVT-BASED ESTIMATION OF RISK CAPITAL AND CONVERGENCE OF HIGH QUANTILES

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

Abstract

We discuss some statistical issues regarding the accuracy of a quantile-based estimation of risk capital. In this context, Extreme Value Theory (EVT) emerges naturally. The paper sheds some further light on the ongoing discussion concerning the use of a semi-parametric approach like EVT and the use of specific parametric models like the g-and-h. In particular, the paper discusses problems and pitfalls evolving from such parametric models when using EVT and highlights the importance of the underlying second order tail behavior.

Keywords: Extreme Value Theory; g-and-h Distribution; Operational Risk; Peaks Over Threshold; Penultimate Approximation; Second Order Regular Variation; Slow Variation; Value-at-Risk

AMS 2000 Subject Classification: Primary 60G70

Secondary 62G32

1. Introduction

Over recent years, we have witnessed a growing interest in the theory and applications of EVT. For instance, textbooks like de Haan and Ferreira [16], Balkema and Embrechts [2] or Resnick [26] discuss new methodological developments together with specific applications to such fields as environmental statistics, telecommunication, insurance and finance. In applying EVT one still runs into theoretical issues which need further study. In this paper we present such a problem, discuss some partial solutions

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

and indicate where more research is needed.

Our starting point is a problem from the regulatory framework (so-called Basel II) of banking and finance. The reader interested in this more applied background is referred to Chapter 1 in McNeil et al. [23]. For the purposes of the present paper, we concentrate on the quantitative modeling of Operational Risk (OR). The latter is defined as *the risk of losses resulting from inadequate or failed internal processes, people and systems or from external events. This definition includes legal risk, but excludes strategic and reputational risk.* OR can be viewed as complementary to the widely studied risk classes Market Risk (MR) and Credit Risk (CR). Without going into a full description of OR data, it suffices to know that risk capital for OR has to be calculated (statistically estimated) using the concept of Value-at-Risk (VaR) at the extreme level of 99.9% and for losses aggregated over a 1-year period; see Section 3 for a definition of VaR. Because of this, early on EVT was recognized as a canonical tool (see Moscadelli [24]) but also criticized for the possible instability of its output, see Dutta and Perry [11]. Degen et al. [10] highlights some of the main problems in applying EVT to OR data and moreover compares and contrasts EVT with the alternative g-and-h approach as championed by Dutta and Perry [11]. One of the main conclusions of these earlier analyses of OR data is that in standard loss severity models used in OR practice, it is the asymptotic behavior of the tail-distribution (in particular the associated slowly varying function in a Pareto-type model) that may cause problems. In the present paper we highlight these issues more in detail. In particular we show that for a simple distribution like the g-and-h, EVT-based estimation for ranges relevant for practice may give answers wich differ significantly from the asymptotics.

In Section 2 we give basic notation, model assumptions and review some standard facts from EVT. Section 3 discusses rates of convergence for quantiles and penultimate approximations. Section 4 looks more carefully into the second order tail behavior under specific model assumptions for EVT applications to OR. We show that the slowly varying function underlying the g-and-h model has second order properties which may give rise to misleading conclusions when such data are analyzed using standard EVT methodology. Section 5 concludes and gives hints for further research.

2. Univariate EVT - background and notation

We assume the reader to be familiar with univariate EVT, as presented for instance in Embrechts et al. [12]. Below we review some basic facts. Throughout we assume that our loss data X are modeled by a continuous df $F(x) = P(X \le x)$ and standardly write $\overline{F} = 1 - F$.

We use the notation $\mathrm{MDA}(H_\xi)$ for Maximum Domain of Attraction of a generalized extreme value df H_ξ; see Embrechts et al. [12] for details. Throughout the paper we restrict our attention to the case $\xi > 0$. Then $F \in \mathrm{MDA}(H_\xi)$ is equivalent to $\overline{F} \in RV_{-1/\xi}$; see for instance Embrechts et al. [12], Theorem 3.3.7. In terms of its tail quantile function $U(x) = F^{\leftarrow}(1 - 1/x)$, this is equivalent to $U \in RV_\xi$. We standardly use the notation $\overline{F} \in RV_{-1/\xi}$ for $\overline{F}(x) = x^{-1/\xi}L(x)$, where L is some slowly varying function in the sense of Karamata, i.e. for all $x > 0$,

$$\lim_{t\to\infty} \frac{L(tx)}{L(t)} = 1. \tag{1}$$

We write L_F and L_U for the slowly varying functions associated with $\overline{F}$ and U respectively. F and U are always assumed to be continuous and sufficiently smooth where needed.

Often it turns ouf to be more convenient to work on a log-log scale for $\overline{F}$ and U. As such, for $\overline{F} \in RV_{-1/\xi}$ with density f we will write

$$\overline{F}(x) = e^{-\Psi(s)} = e^{-s/\xi + \Psi_{L_F}(s)}, \quad s = \log x,$$

where Ψ and Ψ_{L_F}, denote the log-log transform of $\overline{F}$ and L_F respectively; see Appendix A for details. Similarly we define $U(t) = e^{\varphi(r)} = e^{\xi r + \varphi_{L_U}(r)}$, with $r = \log t$. Note that L varies slowly if Ψ_L' vanishes at infinity.

For $F \in \mathrm{MDA}(H_\xi)$ the result below yields a natural approximation for the excess df F_u, defined by $F_u(x) = P(X - u \le x | X > u)$, in terms of the *generalized Pareto distribution (GPD) function* $G_{\xi,\beta}$, defined by $G_{\xi,\beta}(x) = 1 - (1 + \xi x/\beta)^{-1/\xi}$ where $1 + \xi x/\beta > 0$.

Proposition 2.1. (Pickands-Balkema-de Haan.) *For $\xi \in \mathbb{R}$ the following statements are equivalent:*

(i) $F \in \mathrm{MDA}(H_\xi)$,

(ii) There exists a strictly positive measurable function β such that

$$\lim_{u \to x_F} \sup_{x \in (0, x_F - u)} \left| F_u(x) - G_{\xi, \beta(u)}(x) \right| = 0. \tag{2}$$

The defining property of the scale function β is given by the asymptotic relationship $\beta(u) \sim u\xi$, as $u \to \infty$. Moreover, if (2) holds in the case $\xi > 0$ for some function $\beta > 0$, then it also holds for $\beta(u) = \xi u$; see de Haan and Ferreira [16], Theorem 1.2.5.

Remark 2.1. A key property (and hence data assumption) underlying EVT estimation based on Proposition 2.1 is that of *stability*. The class of GPDs is stable under successive conditioning over increasing thresholds, and this with the *same* parameter ξ; see Embrechts et al. [12], Theorem 3.4.13 (c) and Remark 5), p.166. All EVT-based estimation procedures are based on this tail-stability property. It may typically hold for environmental data; whether or not it is tenable for socio-economic data is an issue of current debate.

As a consequence of Proposition 2.1, for $F \in \mathrm{MDA}(H_\xi)$ a natural approximation of the tail $\overline{F}(x)$, for $x \geq u$ and u sufficiently large, is provided by $\overline{F}(u)\overline{G}_{\xi,\beta(u)}(x-u)$. For $\xi > 0$ we may without loss of generality take $\beta(u) = \xi u$ and hence consider the approximation

$$\overline{K}(u,x) := \begin{cases} \overline{F}(x), & x < u, \\ \overline{F}(u)\overline{G}_{\xi,\xi u}(x-u) = c(u)x^{-1/\xi}, & x \geq u, \end{cases} \tag{3}$$

where $c(u) = u^{1/\xi}\overline{F}(u) = L_F(u)$.

For practical purposes, in order to appreciate the goodness of the tail approximation (3) for $x \geq u$, it is important to quantify the rate at which $\overline{K}$ converges to $\overline{F}$, i.e. to determine the rate of convergence in (2)—or equivalently in (1).

In an OR context, rates of convergence for loss severity distributions are discussed in Degen et al. [10]. In the latter paper, the authors focus on the so-called g-and-h distribution recently proposed by Dutta and Perry [11] to model operational risk losses. Recall that a random variable (rv) X is said to have a *g-and-h distribution*, if X satisfies

$$X = a + b\frac{e^{gZ}-1}{g}e^{hZ^2/2}, \quad a, g, h \in \mathbb{R}, b > 0,$$

where $Z \sim N(0,1)$. The linear transformation parameters a and b are of minor importance for our analysis. Unless stated otherwise we therefore restrict our attention to the standard case $a = 0$ and $b = 1$. This class of dfs was introduced in Tukey [28] and studied from a statistical point of view for instance in Hoaglin et al. [18] and Martinez and Iglewicz [22].

Degen et al. [10] show that for $g, h > 0$ (typical for OR data) the g-and-h distribution tail is regularly varying with index $-1/h$, i.e. $\overline{F}(x) = x^{-1/h} L_F(x)$. The corresponding slowly varying L_F is, modulo constants, asymptotically of the form $\exp(\sqrt{\log x})/\sqrt{\log x}$ (see also (7) below) and turns out to be a particularly difficult function to handle from a statistical data analysis point of view. Indeed, below we show that the behavior of L_F in ranges relevant for practice is very different from its limit behavior, which may cause EVT-based estimation methods of h to be very inaccurate over such ranges.

3. Convergence of quantiles

In quantitative risk management, often risk capital charges are based on estimates of high quantiles (Value-at-Risk) of underlying profit-and-loss distributions; see Chapters 1 and 2 in McNeil et al. [23] for details.

Definition 3.1. The *generalized inverse* of a df F,

$$F^{\leftarrow}(q) = \inf\{x \in \mathbb{R} : F(x) \geq q\}, \quad 0 < q < 1,$$

is called the *quantile function* of F. In a financial risk management context, for given q, $F^{\leftarrow}(q)$ is referred to as the $q100\%$ *Value-at-Risk*, denoted by $VaR_q(F)$.

The tail approximation (3) suggests estimating VaR of an (unknown) underlying df $F \in MDA(H_\xi)$, i.e. estimating $F^{\leftarrow}(q)$ for some level $q \in (0,1)$, typically close to 1, by its approximating counterpart $K^{\leftarrow}(u,q)$. By the properties of inverse functions (see for instance Resnick [25], Proposition 0.1) the quantiles $K^{\leftarrow}(.,q)$ converge pointwise to the true value $F^{\leftarrow}(q)$ as $u \to \infty$, but the convergence does not need to be uniform.

Definition 3.2. We say that for some df $F \in \mathrm{MDA}(H_\xi)$, *uniform relative quantile (URQ) convergence* holds if

$$\lim_{u\to\infty} \sup_{q\in(0,1)} \left| \frac{K^{\leftarrow}(u,q)}{F^{\leftarrow}(q)} - 1 \right| = 0;$$

see also Makarov [21], who gives a necessary condition for URQ convergence.

According to Makarov [21], failure of uniform relative quantile (URQ) convergence may lead to unrealistic risk capital estimates. Another possible reason for the discrepancy between an EVT-based methodology and certain parametric approaches for high quantile estimation is provided by the fact that the excess dfs of many loss rvs used in practice (e.g. lognormal, loggamma, g-and-h) show very slow rates of convergence to the GPD. At the model level, this is due to the second order behavior of the underlying slowly varying functions. Consequently, tail index estimation and also quantile (i.e. risk capital) estimation using EVT-based methodology improperly may yield inaccurate results; see Degen et al. [10]. Below we combine both lines of reasoning and embed URQ convergence in the theory of second order regular variation.

3.1. Rates of convergence for quantiles

Assume $\overline{F} \in RV_{-1/\xi}$ for some $\xi > 0$, or equivalently, $U \in RV_\xi$, in terms of its tail quantile function $U(t) = F^{\leftarrow}(1 - 1/t)$. In order to assess the goodness of the tail approximation (3), the rate at which $K^{\leftarrow}(u, q)$ tends to the true quantile $F^{\leftarrow}(q)$ as $u \to \infty$, has to be specified. We are thus interested in the rate at which

$$\frac{U(tx)}{U(t)} - x^\xi \tag{4}$$

tends to 0 as $t \to \infty$.

In the sequel we focus on the rate of convergence in (4) within the framework of the theory of second order regular variation, as presented for instance in de Haan and Stadtmüller [17] or de Haan and Ferreira [16], Section 2.3 and Appendix B.

U is said to be of *second order regular variation*, written $U \in 2RV_{\xi,\rho}$, $\xi > 0, \rho \leq 0$, if for some positive or negative function A with $\lim_{t\to\infty} A(t) = 0$,

$$\lim_{t\to\infty} \frac{\frac{U(tx)}{U(t)} - x^\xi}{A(t)} =: H_{\xi,\rho}(x), \quad x > 0, \tag{5}$$

exists, for some $H_{\xi,\rho}$ which is non-trivial. In that case we necessarily have $H_{\xi,\rho}(x) = x^\xi \frac{x^\rho - 1}{\rho}$ for $x > 0$. Note that for $\xi > 0$, $U \in 2RV_{\xi,\rho}$ is equivalent to $\overline{F} \in 2RV_{-1/\xi,\rho/\xi}$.

In order to find sufficient conditions for (5) to hold, consider the following intuitive reasoning in terms of log-log transforms. Let $U \in RV_\xi$ and assume that U' exists. This ensures that we may write U as

$$U(t) = e^{\varphi(\log t)}, \quad \varphi(r) = \int_1^{e^r} \frac{\varepsilon(y)}{y} dy + c,$$

where $\varepsilon(y) = yU'(y)/U(y)$ and $c = \log U(1)$; see Appendix A. The log-log plot of U depicts the graph of φ. With $s = \log x$ and $r = \log t$ we may write (4) in terms of log-log transforms as

$$\frac{U(tx)}{x^\xi U(t)} - 1 = e^{\varphi(r+s)-\varphi(r)-\xi s} - 1 \sim \varphi(r+s) - \varphi(r) - \xi s, \quad r = \log t \to \infty.$$

The expression $\varphi(r+s)-\varphi(r)-\xi s$ may be approximated by $(\varphi'(r)-\xi)s$ and therefore, the convergence rate of $\frac{U(tx)}{U(t)} - x^\xi$ to 0 is of the same order as the rate at which $\varphi'_{L_U}(\log t) = \varphi'(\log t) - \xi = \frac{tU'(t)}{U(t)} - \xi$ tends to 0 as $t \to \infty$. This motivates the next result.

Theorem 3.1. *Suppose $U(t) = e^{\varphi(\log t)}$ is twice differentiable and $A(t) = \varphi'(\log t) - \xi$. If for some $\xi > 0$ and some $\rho \leq 0$*

i) $\lim_{t\to\infty} \varphi'(t) = \xi$,

ii) $\varphi'(t) - \xi$ *is of constant sign near infinity, and*

iii) $\lim_{t\to\infty} \frac{\varphi''(t)}{\varphi'(t) - \xi} = \rho$,

then, for $x > 0$,

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

with

$$H_{\xi,\rho}(x) = \begin{cases} x^\xi \frac{x^\rho - 1}{\rho}, & \rho < 0, \\ x^\xi \log x, & \rho = 0. \end{cases}$$

Proof. Recall the definition of $A(t) = \varphi'(\log t) - \xi$ and observe that

$$\lim_{t\to\infty} \frac{\frac{U(tx)}{U(t)} - x^\xi}{A(t)} = H_{\xi,\rho}(x) \iff \lim_{r\to\infty} \frac{\varphi(r+s) - \varphi(r) - \xi s}{\varphi'(r) - \xi} = e^{-\xi s} H_{\xi,\rho}(e^s),$$

where $s = \log x$ and $r = \log t$.

By assumption *iii)* we have $\lim_{t\to\infty} tA'(t)/A(t) = \rho$ which, by the Representation Theorem for regularly varying functions (see for instance Bingham et al. [4], Theorem

1.3.1), guarantees $|A| \in RV_\rho$. In particular $A(tx)/A(t) \to x^\rho$ locally uniformly as $t \to \infty$, and we obtain for every $s \in \mathbb{R}$ as $r \to \infty$,

$$\frac{\varphi(r+s)-\varphi(r)-\xi s}{\varphi'(r)-\xi} = \frac{\int_0^s (\varphi'(r+u)-\xi)\,du}{\varphi'(r)-\xi} \quad\longrightarrow\quad \int_0^s e^{\rho u}du = \begin{cases} \frac{e^{\rho s}-1}{\rho}, & \rho<0, \\ s, & \rho = 0. \end{cases}$$

This finishes the proof.

Note that in smooth cases, the rate of convergence is thus uniquely determined by the underlying slowly varying function L_U, i.e. $A(t) = \varphi'_{L_U}(\log t) = tL'_U(t)/L_U(t)$.

Remarks 3.1. (*URQ convergence.*)

i) Define $\Psi(s) = -\log \overline{F}(e^s)$ and $\Psi^u(s) = -\log \overline{K}(u, e^s)$ for $s \in [\log u, \infty)$ (and $\Psi \equiv \Psi^u$ on $(-\infty, \log u)$ by definition of $\overline{K}$). With this notation, URQ convergence holds for F if and only if $\|\Psi^{\leftarrow}(s) - (\Psi^u)^{\leftarrow}(s)\|_\infty \to 0$ on $(-\infty, 0)$ as $u \to \infty$.
It is not difficult to see that URQ convergence holds for F if and only if the log-log transform Ψ_{L_F} of L_F satisfies $\Psi_{L_F}(s) \to c_F$ as $s \to \infty$, for some constant $c_F \in (0,\infty)$.

ii) For U satisfying the assumptions of Theorem 3.1 with $\rho < 0$ we have that $\varphi'(\log x) - \xi \sim Cx^\rho, x \to \infty$, for some constant $C > 0$, and hence $\varphi_{L_U}(t) = \varphi(t) - t\xi$ (or equivalently $\Psi_{L_F}(s) = s/\xi - \Psi(s)$) tends to a finite limit. In particular, a strictly negative second order parameter ρ implies URQ convergence for U (or equivalently F). In the case $\rho = 0$, L_U belongs to the so-called *de Haan class* Π, which is a subclass of the class of the slowly varying functions; see de Haan and Ferreira, Theorem B.3.6. For $L_U \in \Pi$, $\lim_{x\to\infty} L_U(x) =: L_U(\infty)$ exists, but $L_U(\infty)$ but may be infinite; see de Haan and Ferreira [16], Corollary B.2.13.

For $\overline{F} \in RV_{-1/\xi}$, the above theorem allows to calculate the rate at which $K^{\leftarrow}$ tends to the true quantile $F^{\leftarrow}$, or equivalently, the rate at which the corresponding properly scaled excess quantiles $F_u^{\leftarrow}$ converge to $G_{\xi,1}^{\leftarrow}$. For $A(t) = tU'(t)/U(t) - \xi$ satisfying the conditions of Theorem 3.1 we obtain for every $x > 1$,

$$\lim_{u\to\infty} \frac{\left| \frac{F_u^{\leftarrow}(1-1/x)}{\xi u} - G_{\xi,1}^{\leftarrow}(1-1/x) \right|}{A\left(1/\overline{F}(u)\right)} = \frac{1}{\xi} H_{\xi,\rho}(x),$$

where—for F sufficiently smooth—the corresponding convergence rate satisfies

$$A\left(1/\overline{F}(u)\right) = \frac{\overline{F}(u)}{uf(u)} - \xi \sim \xi^2 \frac{uL'_F(u)}{L_F(u)}, \quad u \to \infty.$$

Whereas for distributions with $\rho < 0$, this convergence is rather fast, it may be very slow in the case $\rho = 0$. The loggamma for instance is well-known for its slow convergence properties with $A(1/\overline{F}(u)) = O\left(1/\log u\right)$. The situation for the g-and-h is even worse with $A(1/\overline{F}(u)) = O\left(1/\sqrt{\log u}\right)$.

Summing up, in terms of the second order parameter ρ, the tail $\overline{F}$ (or equivalently U) will be "better" behaved if $\rho < 0$, than in the case $\rho = 0$. In the former case the convergence in (5) is not "too slow" in the sense that the rate function $|A|$ is regularly varying with index $\rho < 0$, i.e. if the influence of the nuisance term L_U, or equivalently L_F, vanishes fast enough in terms of L_U (or L_F) behaving "nicely" and tending to some positive constant (thus implying URQ convergence for F). In the case $\rho = 0$, the rate function $|A|$ is slowly varying and hence the excess quantiles typically converge very slowly. However, in certain cases the slow convergence rate may be improved using the concept of penultimate approximation.

3.2. Penultimate approximations

So far we have been concerned with the *ultimate approximation*, i.e. for every $x > 0$ and for large values of t,

$$U(tx) \approx x^{\xi} U(t).$$

One method to improve the rate of convergence in the above approximation goes back to the seminal work of Fisher and Tippett [13]. More recent accounts on this are found for instance in Cohen [7], Gomes and de Haan [14] or Worms [29].

The basic idea behind *penultimate approximations* is to vary the shape parameter ξ as a function of the threshold t, i.e. to consider

$$U(tx) \approx x^{\xi(t)} U(t),$$

with $\xi(t) \to \xi$ for $t \to \infty$, where one hopes to improve the convergence rate by choosing $\xi(.)$ in an appropriate way.

In order to illustrate how to find a feasible function $\xi(.)$ consider the following. In terms of log-log transforms, the ultimate approximations uses that, for large values of

r,

$$\varphi(r+s) \approx \varphi(r) + \xi s,$$

where $r = \log t$ and $s = \log x$, i.e. φ is approximated linearly by a straight line with slope ξ. A better approximation might be achieved, if for large values of r, φ is still approximated linearly, but now by its tangent line in the respective points $r = \log t$, i.e. by a straight line with slope φ', leading to

$$\varphi(r+s) \approx \varphi(r) + \varphi'(r)s.$$

Thus, a reasonable choice of a threshold-dependent shape parameter is $\xi(t) = \varphi'(\log t) = tU'(t)U(t)$.

At this point it is worth noting that there is a close connection between the theory of penultimate approximations and the theory of second order regular variation. Suppose U satisfies the conditions given in Theorem 3.1. In that case we obtain for large values of t,

$$U(tx) \approx U(t) \left(x^{\xi} + A(t) H_{\xi,\rho}(x) \right),$$

or equivalently, for large values of $r = \log t$,

$$\varphi(r+s) \approx \varphi(r) + \xi s + \left(\varphi'(r) - \xi\right) e^{-\xi s} H_{\xi,\rho}\left(e^{s}\right),$$

which, by definition of $H_{\xi,\rho}$ for $\rho = 0$, is the same as $\varphi(r+s) \approx \varphi(r) + \varphi'(r)s$. As a consequence, in the case $\rho = 0$, the theory of second order regular variation yields (*asymptotically*) the same approximation as the penultimate theory. Note however, that we may easily construct examples where the second order theory does not apply but the penultimate does.

Example 3.1. Consider $U(t) = e^{\varphi(\log t)}$, with $\varphi(r) = \xi r + \varphi_{L_U}(r) = \xi r + \sin\sqrt{r}$. In this case, φ_{L_U} changes sign infinitely often as we move out and thus Theorem 3.1 does not apply. A penultimate approximation $\varphi(r+s) = \varphi(r) + \varphi'(r)s$ may nevertheless be considered. In particular, the approximation error $e_r(s) = \varphi(r+s) - \varphi(r) - \varphi'(r)s$ is of the order $O(s^2)$ for $s \to 0$, whereas the error in the ultimate case $\varphi(r+s) - \varphi(r) - \xi s$ is of the order $O(s)$.

The above example shows that, although no second order improvement exists in that particular case, the penultimate approximation may *locally* still lead to an improvement over the ultimate approximation. In addition, we shall show below that in the case $\rho = 0$, the rate of convergence in the penultimate approximation may indeed (asymptotically) improve compared to the ultimate approximation.

Intuitively it is clear that the rate at which $\frac{U(tx)}{U(t)} - x^{\xi(t)}$ tends to 0 is of the same order as the rate at which the linear approximations (i.e. the tangent lines in points t) approach the straight line with slope ξ. Hence, the speed at which the penultimate convergence rate $a(.)$ tends to 0 is of the same order as the speed at which the slope φ' tends to its ultimate value ξ which is measured by φ''. So as a candidate for the penultimate convergence rate we choose $a(t) = \varphi''(\log t) = t\xi'(t)$. Condition iii) of Theorem 3.1, $\lim_{t\to\infty} \frac{\varphi''(t)}{\varphi'(t)-\xi} = \rho$ for some $\rho \leq 0$, implies that for this choice of a, the convergence rate may asymptotically only be improved in cases where $\rho = 0$ and we have $a(t) = o(A(t))$ for $t \to \infty$. Indeed, under the conditions discussed above and under the following additional condition

$$\frac{\varphi'''(x)}{\varphi''(x)} \to 0 \text{ or equivalently } \frac{t\xi''(t)}{\xi'(t)} \to -1, \quad x = \log t \to \infty,$$

the (improved) penultimate rate of convergence may be given as follows.

Theorem 3.2. *Let U satisfy the conditions of Theorem 3.1 with $\rho = 0$ and define $\xi(t) = tU'(t)/U(t) = A(t) + \xi$. If $\lim_{t\to\infty} \frac{t\xi''(t)}{\xi'(t)} = -1$, and ξ' is of constant sign near infinity, then, for all $x > 0$,*

$$\lim_{t\to\infty} \frac{\frac{U(tx)}{U(t)} - x^{\xi(t)}}{tA'(t)} = J_\xi(x),$$

where

$$J_\xi(x) = \frac{1}{2}x^\xi (\log x)^2 .$$

Proof. Following the lines of the proof of Theorem 3.1 note first that for $s = \log x$ and $r = \log t$,

$$\lim_{t\to\infty} \frac{\frac{U(tx)}{U(t)} - x^{\xi(t)}}{tA'(t)} = J_\xi(x) \Longleftrightarrow \lim_{r\to\infty} \frac{\varphi(r+s) - \varphi(r) - \varphi'(r)s}{\varphi''(r)} = \frac{1}{2}s^2.$$

For every $s \in \mathbb{R}$ we have

$$\frac{\varphi(r+s) - \varphi(r) - \varphi'(r)s}{\varphi''(r)} = \int_0^s \int_0^y \frac{\varphi''(r+z)}{\varphi''(r)}\, dzdy.$$

Moreover, by assumption $t\xi''(t)/\xi'(t) \to -1$ (or equivalently $\varphi'''(x)/\varphi''(x) \to 0$) which implies that $|\varphi''| \in RV_0$ which guarantees $\varphi''(r+s)/\varphi''(r) \to 1$ locally uniformly for $r \to \infty$, and hence for $r \to \infty$,

$$\int_0^s \int_0^y \frac{\varphi''(r+z)}{\varphi''(r)} \, dzdy \to \int_0^s \int_0^y 1 \, dzdy = \frac{1}{2}s^2,$$

which finishes the proof.

Remarks 3.2. i) We want to stress again that is important to distinguish between the second order theory which is an *asymptotic* theory (i.e. is concerned with the limit behavior) while the penultimate theory is a *lcoal* theory. Only for the special case $\rho = 0$ and under certain additional conditions, the second order theory and the penultimate theory yield *asymptotically* the same approximation.

ii) From the proof of Theorem 3.2 it is clear that, although the original rate is improved *asymptotically*, i.e. $tA'(t) = o\left(A(t)\right)$ for $t \to \infty$, the improvement is not spectacular as the new rate $tA'(t)$ is again slowly varying. Nevertheless, *locally* the improvements may be considerable as we show in the next paragraph.

3.3. Implications for practice

To illustrate the above results, we compare the (theoretical) relative approximation error $e(u) := \left|\frac{K^{\leftarrow}(u,1-1/x)}{F^{\leftarrow}(1-1/x)} - 1\right|$ for the 99.9% quantiles (confidence level required under Basel II) as a function of the threshold u, in the ultimate and penultimate approximation for certain frequently OR-loss severity models. Besides the well-known Burr and loggamma distribution (see for instance Embrechts et al. [12], Chapter 1), we consider the g-and-h, the modified Champernowe and the generalized Beta distribution.

The modified Champernowne was recently proposed by Buch-Larssen et al. [5] in an OR context. Its tail df is given by

$$\overline{F}(x) = \frac{(M+c)^{\alpha} - c^{\alpha}}{(x+c)^{\alpha} + (M+c)^{\alpha} - 2c^{\alpha}}, \quad \alpha, M, x > 0, c \geq 0,$$

hence $\overline{F} \in RV_{-\alpha}$. The density of a generalized Beta (GB2) distribution is given by

$$f(x) \propto \frac{x^{ap-1}}{\left(1 + \left(\frac{x}{b}\right)^{a}\right)^{p+q}}, \quad a, b, p, q, x > 0,$$

so that $f \in RV_{-aq-1}$; see for instance Dutta and Perry [11] for its use in OR.

In Figure 1 we show the approximation errors $e(.)$ in % for the g-and-h and the loggamma distribution (left panel) and for the Burr, the modified Champernowne and the GB2 distribution (right panel). To enable a qualitative comparison across different distributions we take the thresholds as quantile levels q and scale the horizontal axis by the 99.9% quantile.

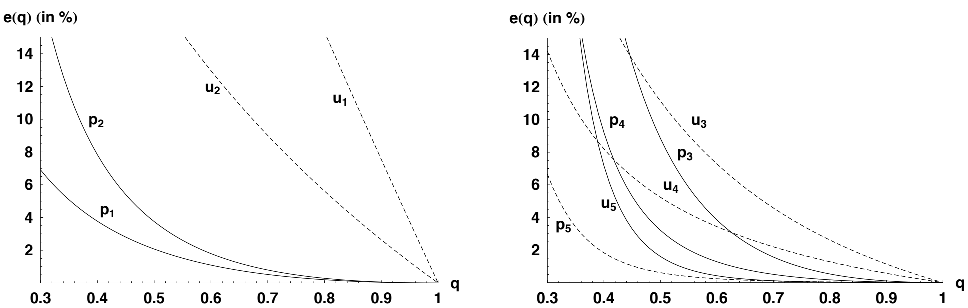

FIGURE 1: Approximation error for the 99.9% quantile of the g-and-h ($g = 2,\ h = 0.2$; labels $u1$, $p1$), the loggamma ($\alpha = 5,\ \beta = 5$; labels $u2$, $p2$), the Burr ($\alpha = 3, \tau = 1, \kappa = 5/3$, labels $u3$, $p3$), the modified Champernowne ($\alpha = 5, M = 1, c = 1/4$; labels $u4$, $p4$) and the GB2 distribution ($a = 5, b = 1, p = 1, q = 1$; labels $u5$, $p5$).

In order to compare quantitatively and to check how the GPD-approximation for high quantiles performs, we fix a relative error level of $e^{99.9\%}(u) = 5\%$, say, and compute the excess probabilities over the corresponding u levels. In practice, in order to estimate a 99.9% quantile using the POT-method, a certain amount N_u of data exceeding this threshold u is needed (we take u such that $e^{99.9\%}(u) = 5\%$), so as to come up with reasonable estimates. For illustrative purposes we choose N_u to be 100. From this we may infer the number n of data points we would expect to have to generate, in order to have $N_u = 100$ excesses over the threshold u for a given relative error $e^{99.9\%}(u) = 5\%$; see Table 1 for the results.

From the five loss models considered above, the Burr, the modified Champernowne and the GB2 satisfy $\rho < 0$ whereas the loggamma and the g-and-h distribution have second order parameter $\rho = 0$. In the latter case, n increases vastly to about 47780 and 91540 for the loggamma and the g-and-h, respectively, reflecting the slow convergence properties of these distributions.

TABLE 1: Expected number of data points n needed to get $N_u = 100$ exceedances over a fixed threshold u for the distributions of Figure 1 with the respectively specified parameter values.

	threshold u s.t. $e^{99.9\%}(u) = 5\%$	n
g-and-h (u_1)	585	91540
g-and-h (p_1)	221	27740
Loggamma (u_2)	15.6	47780
Loggamma (p_2)	8.87	6260
Burr (u_3)	2.55	19070
Burr (p_3)	2.13	9190
Modified Champernowne (u_4)	2.41	4480
Modified Champernowne (p_4)	2.15	2720
Generalized Beta (u_5)	1.28	440
Generalized Beta (p_5)	1.68	1440

Using penultimate approximations above a reasonably high threshold, the number of data needed to achieve the same level of accuracy may be lowered significantly in all cases but for the GB2. Note however that in the GB2 case the convergence is rather fast and the approximation error $e(u)$ is negligible for high threshold values u anyway. Though the theory of penultimate approximations seems to be very promising towards the (theoretical) improvement of high-quantile estimation accuracy, its practical relevance may be limited since the slope φ' has to be estimated from the data. More statistical work would be highly useful here.

Remark 3.3. The above examples are of course idealized since we assume the underlying distributions F to be known and hence also the corresponding tail index ξ of the GPD G_ξ. In practice we will encounter an additional error source due to estimation errors of the parameters.

From an applied risk management point of view our analysis admits the following conclusion. The closer the second order parameter ρ is to zero, the slower EVT-based estimation techniques converge. Thus, if data seem to be modeled well by a df F with second order parameter $\rho = 0$, the amount of data needed in order to come up with reasonable results may be prohibitively large. In an operational risk context this is also one of the main reasons why banks have to combine internal loss-data with external data and expert opinion, leading to further important statistical issues; see Lambrigger et al. [20]. In addition, for distributions with "bad" second order behavior (i.e. $\rho = 0$) the exact shape of the associated slowly varying function L_F may deteriorate the situation further and give rise to misleading conclusions about the underlying data. As we already saw, a prime example of this situation, important for practice, is provided by the g-and-h slowly varying function ($g, h > 0$), which is analyzed in more detail in the next section.

4. A slowly varying function with a sting in its tail

According to Dutta and Perry [11], the typical parameter values of a g-and-h df F used to model OR loss severities are in a range around $g \in (1.7, 2.3)$ and $h \in (0.1, 0.4)$. In the sequel we therefore always assume g and h to be strictly positive, hence $\overline{F} \in RV_{-1/h}$. Adopting the notation of Degen et al. [10], we consider $X \sim$ g-and-h with df $F(x) = \Phi(k^{-1}(x))$, where $k(x) = \frac{e^{gx}-1}{g} e^{hx^2/2}$ and Φ denotes the standard normal df. In Figure 2 we plot $\overline{F}$ on a log-log scale for OR-typical parameter values. As $\overline{F} \in RV_{-1/h}$, a straight line with slope $-1/h$ is to be expected as we move out in the right tail.

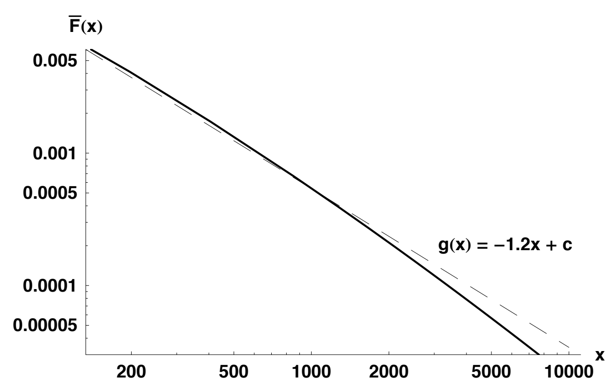

FIGURE 2: Log-log plot of the tail of a g-and-h distribution with $g = 2$ and $h = 0.2$ and straight line $g(.)$ as a reference line with slope -1.2.

According to Figure 2, the log-log plot is almost linear over a large region of practical interest (for quantile levels of 90% up to 99.99%). Therefore, over such ranges, the g-and-h tail behavior is close to an exact Pareto and thus the influence of the slowly varying part seems to be minimal. Figure 2 is fallacious however, as the slope of a linear approximation $g(.)$ implies a tail index of around 0.8 whereas the theoretical tail index was chosen to be $h = 0.2$. The consequences for statistical estimation of this may be better understood by the concept of what we will call *local heavy-tailedness*.

4.1. Local heavy-tailedness

Consider $\overline{F}(x) = x^{-1/\xi} L_F(x) \in RV_{-1/\xi}$, $\xi > 0$, where F has a density f. Recall that we then may write $\overline{F}$ as

$$\overline{F}(x) = e^{-\Psi(s)} = e^{-s/\xi + \Psi_{L_F}(s)}, \quad s = \log x,$$

with Ψ and Ψ_{L_F} denoting the log-log transforms of $\overline{F}$ and L_F respectively; see Appendix A. As a graph of $-\Psi$ corresponds to the log-log plot of $\overline{F}$, the total amount of heavy-tailedness at a point x is measured by $\Psi'(\log x)$. It consists of the ultimate heavy-tailedness of the model (tail index ξ) and an additional source of local heavy-tailedness due to L_F. The local heavy-tailedness is measured by the slope of $\log L_F$ which is given by

$$\Psi'_{L_F}(s) = \frac{1}{\xi} - \Psi'(s) = \frac{1}{\xi} - \frac{xf(x)}{\overline{F}(x)}, \quad s = \log x. \tag{6}$$

Clearly, depending on the underlying model F, the amount of local heavy-tailedness due to the shape of its associated slowly varying function L_F may be significant. As indicated above, this is particularly evident in the case of the g-and-h slowly varying function, for which it turns out that the behavior of $\overline{F}$ (or L_F) in ranges relevant for risk management applications is very different from its ultimate asymptotic behavior. Neglecting this issue for data of (or close to) g-and-h type may lead to problems when applying standard EVT methodology. Whereas this issue is well known from a theoretical point of view within the EVT community (see for instance Resnick [25], Exercise 2.4.7), it is somewhat surprising that it manifests itself so clearly in a fairly straightforward, and increasingly used parametric model like the g-and-h.

To get a feeling for the behavior of the g-and-h slowly varying function L_F we show in Figure 3 a log-log plot of L_F (left panel) with corresponding slope (right panel) for

g and h parameter values typical for OR (the function k is inverted numerically using a Newton algorithm with an error tolerance level of 10^{-13}).

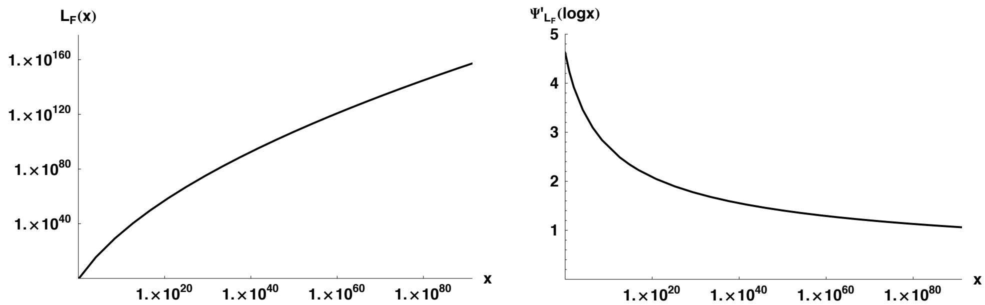

FIGURE 3: Log-log plot of the g-and-h slowly varying function L_F with $g = 2$ and $h = 0.2$ (left panel) with corresponding slope (right panel) as defined in (6).

The almost linear log-log plot suggests that the function L_F behaves approximately like a power function $x^{1/\eta}$ for some $\eta > 0$ and with $1/\eta$ given in the slope plot. The asymptotic behavior of the log-log transform of $\overline{F}$ is given by

$$-\Psi(s) = -\frac{1}{h}s + \frac{\sqrt{2}g}{h^{3/2}}\sqrt{s} - \frac{1}{2}\log s - c + O\left(\frac{1}{\sqrt{s}}\right), \quad s = \log x \to \infty, \tag{7}$$

with $c = \frac{1}{2}\log\frac{4\pi}{h} + \frac{g^2}{h^2} + \frac{\log g}{h}$; see Appendix A, equation (11).

The deviation from exact power-law decay (i.e. the deviation from linearity in a log-log plot) is due to the slowly varying part. The amount of local heavy-tailedness is measured by the slope Ψ'_{L_F} of $\log L_F$, which behaves like

$$\Psi'_{L_F}(s) = \frac{g}{\sqrt{2}h^{3/2}}\frac{1}{\sqrt{s}} - \frac{1}{2s} + O\left(\frac{1}{s^{3/2}}\right), \quad s = \log x \to \infty. \tag{8}$$

Therefore, the rate at which the influence of L_F vanishes, i.e. the rate at which Ψ'_{L_F} tends to 0, is of the order $O\left(1/\sqrt{\log x}\right)$, $x \to \infty$.

Equation (8) gives us a first impression of how unpleasant the g-and-h slowly varying function might be. Indeed, its slow convergence properties together with its (power-like) behavior in ranges relevant for OR practice may lead to serious difficulties in the statistical estimation of extremes based on EVT, *given that data follow such a model.* At this point we like to stress that this is *not* a weak point of EVT but should more be viewed as a warning against "gormless guessing" of a parametric model, and this in the words of Richard Smith and Jonathan Tawn; see Embrechts et al. [12], Preface p.VII. In the next paragraph we study EVT estimation *within* a g-and-h model, i.e. for g-and-h generated data.

4.2. Tail index estimation for g-and-h data

In Degen et al. [10], the problem of the estimation of the tail index h within a g-and-h model was pointed out using the Hill estimator. To emphasize that EVT-based tail index estimation for g-and-h data may be problematic whatever method one uses, below we shall work with the increasingly popular POT-MLE method for which the statistical basis was laid in the fundamental papers of Davison [8], Smith [27] and Davison and Smith [9]. For further background reading, see Embrechts et al. [12]. We additionally implemented other tail index estimators like the moment estimator, a bias-reduced MLE or an estimator based on an exponential regression model as discussed in Beirlant et al. [3]. As can be expected form the discussion in the previous paragraphs, all these estimators led to similar conclusions and we therefore refrain from showing those results.

In Figure 4 we simulated $n = 10^4$ observations from a g-and-h model ($g = 2, h = 0.2$) and plot the POT-MLE of the tail index h as a function of the number of exceedances used, together with the 95% confidence bounds.

At first glance, Figure 4 suggests that the df of the underlying data follows nearly perfectly an exact Pareto law or at least converges rather fast towards an exact Pareto law. Indeed, the deviation from exact power law behavior seems to vanish quickly as the MLE behaves stable and is flat over a large region of thresholds (as was of course to be expected from Figure 2). As a consequence one would accept an estimate of the tail index of around $\widehat{h} \approx 0.85$ (compare this value to the 0.8 implied by Figure 2). EVT-based estimation thus significantly overestimates the true parameter $h = 0.2$, suggesting a rather heavy tailed model ($\widehat{\alpha} = 1/\widehat{h} \approx 1.2$, i.e. infinite variance), whereas the data are simulated from a model with $1/h = 5$, which has moments finite up to order five. The reason behind this lies in the behavior of the underlying slowly varying function over moderate ranges. More precisely, for g-and-h data ($g = 2$, $h = 0.2$) with for practice reasonable sample sizes, a number of $k = 500$, say, largest order statistics is

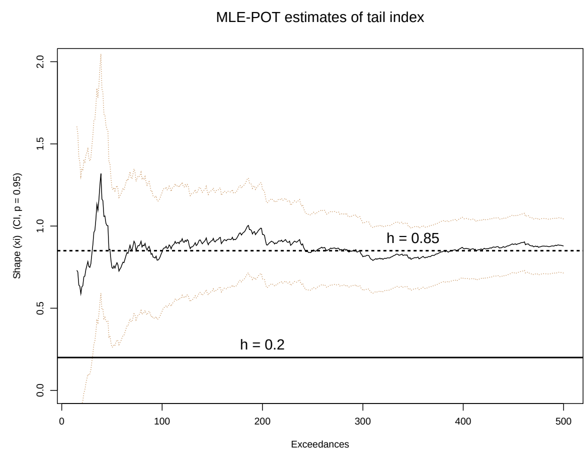

FIGURE 4: Theoretical tail index (solid line) and MLE-POT estimates of $\xi = h$ for g-and-h data with $g = 2$ and $h = 0.2$, based on $n = 10^4$ observations.

taken into account for the estimation of h. For this realistic choice of parameter values of g and h, the values of such order statistics will typically range from around 10 to around 10^3. Over such ranges the slope of Ψ_{L_F} is nearly constant and hence L_F grows approximately like a power $x^{1/\eta}$ with (averaged) $1/\eta \approx 3.8$; see Figure 5. Therefore, in ranges relevant for practical applications, the "regularly varying"-like slowly varying part L_F adds a significant amount of local heavy-tailedness to the model. Together with the asymptotic tail decay of $\overline{F} \in RV_{-5}$, the local power-like growth of L_F leads to a regular variation index of around -1.2, i.e. to a "local" tail index h_{loc} of $\overline{F}$ of around 0.83.

Remark 4.1. The notion of regular variation is an asymptotic concept, hence our emphasis on the distinction between local (i.e. finite ranges relevant for practice) and asymptotic tail behavior. Due to its slow convergence properties, the (asymptotic) tail index for the g-and-h is significantly overestimated with standard (i.e. ultimate) EVT estimation methods. On the other hand, after a penultimate correction accounting for the local heavy-tailedness due to the behavior of the slowly varying part, the behavior of $\overline{F}$ over ranges relevant for practice is captured well by EVT methods (the MLE in Figure 4 is rather stable and flat). This is due to the extremely slow decay of Ψ'_{L_F},

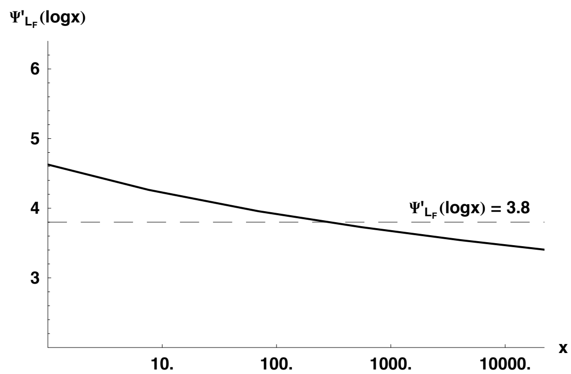

FIGURE 5: Slope of the log-log plot of the g-and-h slowly varying function ($g = 2, h = 0.2$) in a range relevant for operational risk.

which is close to constant over large ranges. Over such ranges, the g-and-h therefore behaves approximately like an exact Pareto, for which EVT-based estimation methods are known to perform well.

Furthermore, as can be seen from (8), the accuracy of tail index estimation crucially depends on the ratio $g/h^{3/2}$ of the g-and-h parameters. For parameter values of g and h relevant for operational risk applications (see Dutta and Perry [11], p.43), $g/h^{3/2}$ is relatively large and thus the slope of Ψ_{L_F}—the share of additional local heavy-tailedness—is large.

For cases with $g/h^{3/2}$ small, for instance $g = 0.1$ and $h = 0.5$, Ψ'_{L_F} is less than 0.2 over a large range, which includes a major part of largest order statistics of reasonable sample sizes, and thus the influence of L_F is minor, i.e. the difference between local and asymptotic behavior is marginal. Consequently the accuracy of EVT-based tail index estimation increases considerably and the estimate captures the structure of the data correctly; see for example Figure 6.

4.3. Risk capital estimation for operational risk

Based on the 2004 loss data collection exercise (LDCE) for operational risk data, Dutta and Perry [11] note that in the estimation of OR risk capital (1-year 99.9% VaR) there seems to be a serious discrepancy between an EVT-based approach and the g-and-h approach. They find that EVT often yields unreasonably high risk capital estimates

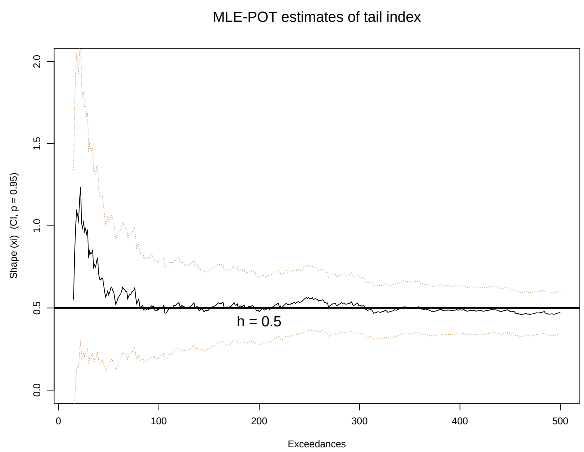

FIGURE 6: Theoretical tail index (solid line) and MLE-POT estimates of $\xi = h$ for g-and-h data with $g = 0.1$ and $h = 0.5$, based on $n = 10^4$ observations.

whereas the g-and-h approach would not. Fitting a g-and-h model to the 2004 LDCE data (aggregated at enterprise level), Dutta and Perry [11] find $\xi = h \in (0.1, 0.4)$ for different banks. In contrast to that, applying EVT methodology to the 2002 LDCE data (aggregated at business line level), Moscadelli [24] comes up with infinite mean models (i.e. $\xi > 1$) for 6 out of 8 business lines. Based on this, resulting risk capital estimates may be expected to differ widely for the two approaches. See also Jobst [19] on this issue; in that paper it is claimed that for typical OR data the g-and-h approach is superior only from a confidence level of 99.99% onwards.

As we do not possess the data underlying these analyses, our findings below, based on simulated data, may be academic in nature. However, even though EVT-based techniques may, for simulated g-and-h data, lead to completely wrong estimates of the (asymptotic) tail index h (see Figure 4), this does not need to carry over to estimates of relevant risk measures like VaR or return periods. Indeed, based on $n = 10^4$ simulated g-and-h data ($g = 2, h = 0.2$), the MLE-POT estimates of the 99.9% quantile seem to be rather accurate (though the 95% confidence band is quite broad); see Figure 7.

The seeming incompatibility of the two Figures 4 and 7 can be explained by taking the effect of local heavy-tailedness caused by the slowly varying part into account. For $g = 2$ and $h = 0.2$, the theoretical 99.9% g-and-h quantile is approximately 626. The

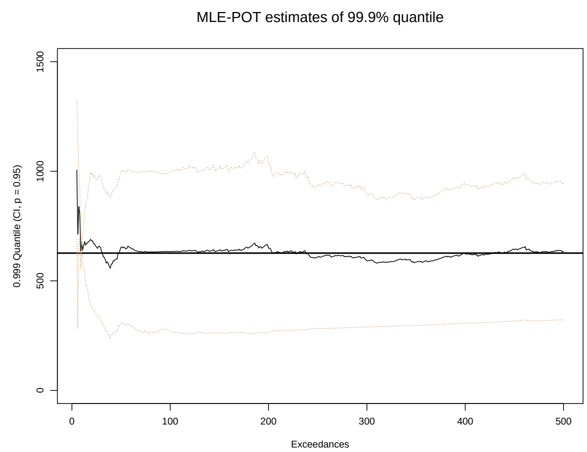

Figure 7: Theoretical 99.9% g-and-h quantile (solid line) and MLE-POT estimates for $g = 2$ and $h = 0.2$, based on $n = 10^4$ observations.

amount of local heavy-tailedness in this point is $\Psi'_{L_F}(626) \approx 3.7$. Moreover, recall that over large ranges the g-and-h behaves approximately like an exact Pareto (with tail index $\neq h$) and thus the situation is as if we were to estimate the 99.9% quantile of a distribution which seems to be modeled well by an exact Pareto distribution with parameter $\alpha = -5 + 3.7 = -1.3$ (i.e. with tail index $\xi \approx 0.77$). This quantile is then estimated using $k = 500$, say, upper order statistics, whose values will—for a sample of $n = 10^4$ observations—typically be ranging from around 10 to around 10^3. Due to the almost constant slope of Ψ_{L_F} (≈ 3.8 averaged) over this range, the 500 order statistics thus seem to be modeled well by an exact Pareto distribution with parameter $\alpha = -1.2$ (i.e. $\xi \approx 0.83$). This may explain the good performance of the MLE-POT estimates of the 99.9% quantile in Figure 7.

Clearly, these quantile estimates are likely to get worse in cases where one is estimating far out-of-sample quantiles. Especially in an OR context, estimating at a level of 99.9% is a serious issue as today's OR loss databases are rather sparse. But still, as the influence of L_F changes only very slowly over large ranges (we are in the $\rho = 0$ case), the size of estimation errors is not nearly as large as in the case for tail index estimation. In this sense, the $\rho = 0$ case for high quantile estimation based on EVT is not necessarily as troublesome as can be expected from the well-known poor performance of tail-index estimation in such a case. A more in depth study of this phenomenon would however be highly desirable and have important consequences for QRM-practice.

5. Conclusion

In this paper we highlight some issues regarding a quantile-based estimation of risk capital motivated by the Basel II regulatory framework for Operational Risk. From a theoretical point of view, EVT-based estimation methodologies of high quantiles arise very naturally. Our main results are as follows.

First, according to Makarov [21], failure of URQ convergence may lead to inaccurate risk capital estimates. We complement these findings by showing that for sufficiently smooth $\overline{F} \in RV_{-1/\xi}$, $\xi > 0$, the asymptotic behavior of the associated slowly varying function L_F determines whether or not URQ convergence holds. This then allows to embed URQ convergence in the framework of second order regular variation for quantiles; $L_F(x) \to c \in (0, \infty)$ implies a second order parameter $\rho < 0$, whereas $L_F(x) \to \infty$ (or 0) implies $\rho = 0$. In the latter case, the slow convergence properties together with the possibly delusive behavior of L_F may cause serious problems when applying standard EVT methodology.

Second, we stress the fact that, when using EVT methodology, the second order behavior of the underlying distribution, which (in smooth cases) is fully governed by its associated slowly varying function, is crucial. If data are well modeled by a distribution with "bad" second order behavior, i.e. with second order parameter $\rho = 0$, EVT-based estimation techniques will typically converge slowly. As a consequence, the amount of data needed in order for EVT to deliver reasonable results may be unrealistically high and largely incompatible with today's situation for OR data bases. The idea of penultimate approximations seems very promising in this respect. So far this concept has been of a more theoretical nature and further applied research would be desirable.

Third, the g-and-h distribution of Dutta and Perry [11] corresponds to a class of loss dfs for which the slowly varying function L_F $(g, h > 0)$ is particularly difficult to handle. Due to its slow convergence properties ($\rho = 0$), its behavior in ranges relevant for OR practice is very different from its ultimate asymptotic behavior. For broad ranges of the underlying loss values, the slowly varying function L_F behaves like a regularly varying function, putting locally some extra weight to the tail $\overline{F}$. As a consequence, standard EVT-based tail-index estimation (asymptotic behavior matters) may result in completely wrong estimates. However, this poor performance does not need to carry over to high quantile estimation (finite range behavior matters).

For risk management applications in general and operational risk in particular, a key property to look for is the second order behavior of the underlying loss severity models. Models encountered in practice often correspond to the case $\rho = 0$. Especially in the latter case, more research on the statistical estimation of high quantiles using EVT is needed; see for instance Gomes and Pestana [15] and references therein for some ideas.

Finally: as already discussed in Remark 2.1, EVT assumes certain tail-stability properties of the underlying loss-data. These may or may not hold. In various fields of application outside of finance and economics these properties seem tenable, and hence EVT has established itself as a most useful statistical modeling tool. Within financial risk management, discussions on stability are still ongoing and may lie at the basis of critical statements on the use of EVT; see for instance Christoffersen et al. [6], Dutta and Perry [11] and Jobst [19]. In their discussion on the forecasting of extreme events, the authors of the former paper state "Thus, we believe that best-practice applications of EVT to financial risk management will benefit from awareness of its limitations, as well as the strengths. When the smoke clears, the contribution of EVT remains basic and useful: it helps us to draw smooth curves through the extreme tails of empirical survival functions in a way that is consistent with powerful theory. Our point is simply that we should not ask more of the theory than it can deliver."

We very much hope that our paper has helped in lifting a bit of the smoke screen and will challenge EVT experts to consider more in detail some of the statistical challenges related to the modeling of extremes in financial risk management.

6. Acknowledgements

The authors would like to thank Guus Balkema for his many valuable comments and suggestions that helped considerably to improve on previous versions of this paper and Johanna Nešlehová for early discussions related to the project of uniform quantile

convergence.

References

[1] Abramowitz, M. and Stegun, I. A. (1972). *Handbook of Mathematical Functions.* Dover Publications, New York.

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

[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] Buch-Larssen, T., Nielsen, J. P., Guillén, M. and Bolancé, C. (2005). Kernel density estimation for heavy-tailed distributions using the Champernowne transformation. *Statistics* **39,** 503–518.

[6] Christoffersen, P. F., Diebold, F. X. and Schuermann, T. (1998). Horizon problems and extreme events in financial risk management. *Economic Policy Review, Federal Reserve Bank of New York* 109–118.

[7] Cohen, J. P. (1982). Convergence rates for the ultimate and penultimate approximations in extreme-value theory. *Advances in Applied Probability* **14,** 833–854.

[8] Davison, A. C. (1984). Modelling excesses over high thresholds, with an application. In *Statistical Extremes and Applications.* ed. J. Tiego de Oliveira. Reidel, Dordrecht pp. 461–482.

[9] Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds (with discussion). *J. Roy. Statist. Soc. B* **52,** 393–442.

[10] Degen, M., Embrechts, P. and Lambrigger, D. D. (2007). The quantitative modeling of operational risk - between g-and-h and EVT. *ASTIN Bulletin* **37,** 265–291.

[11] Dutta, K. and Perry, J. (2006). A tale of tails: an empirical analysis of loss distribution models for estimating operational risk capital. Federal Reserve Bank of Boston, Working Paper No 06-13.

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

[13] 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.

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

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

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

[17] Haan, L. de and Stadtmüller, U. (1996). Generalized regular variation of second order. *J. Austral. Math. Soc.* **61,** 381–395.

[18] Hoaglin, D. C., Mosteller, F. and Tukey, J. W. (1985). *Exploring Data Tables, Trends, and Shapes.* Wiley, New York.

[19] Jobst, A. A. (2007). Operational risk: the sting is still in the tail but the poison depends on the dose. *Journal of Operational Risk* **2(2),** 3–59.

[20] Lambrigger, D. D., Shevchenko, P. and Wüthrich, M. (2007). The quantification of operational risk using internal data, relevant external data and expert opinions. *Journal of Operational Risk* **2(3),** 3–27.

[21] Makarov, M. (2006). Extreme value theory and high quantile convergence. *Journal of Operational Risk* **1(2),** 51–57.

[22] Martinez, J. and Iglewicz, B. (1984). Some properties of the Tukey g and h familiy of distributions. *Communications in Statistics - Theory and Methods* **13,** 353–369.

[23] McNeil, A. J., Frey, R. and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques, Tools.* Princeton University Press, Princeton.

[24] 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.

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

[26] Resnick, S. I. (2007). *Heavy-tail Phenomena: Probabilistic and Statistical Modeling.* Springer, New York.

[27] Smith, R. L. (1987). Estimating tails of probability distributions. *Ann. Statist.* **15,** 1174–1207.

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

[29] Worms, R. (2002). Penultimate approximation for the distribution of the excesses. ESAIM: Probability & Statistics **6,** 21–31.

Appendix A.

By our standing assumption F and $F^{\leftarrow}$ are throughout assumed to be continuous. We shall furthermore always assume sufficient smoothness for F and $F^{\leftarrow}$ where necessary. Consider $\overline{F}(x) = x^{-1/\xi} L_F(x) \in RV_{-1/\xi}$, $\xi > 0$, and denote by f the density of F. This is equivalent to assuming $\overline{F}$ to be *normalized* regularly varying with index $-1/\xi$ (see Bingham et al. (1987), Section 1.3) and ensures that we may write $\overline{F}$ as

$$\overline{F}(x) = e^{-\Psi(\log x)}, \quad \Psi(s) = \int_1^{e^s} \frac{\eta(y)}{y} dy - c, \tag{9}$$

with $\eta(y) = yf(y)/\overline{F}(y)$ and $c = \log \overline{F}(1)$. So the graph of $-\Psi$ corresponds to the log-log plot of $\overline{F}$ and we have $\Psi'(s) = xf(x)/\overline{F}(x)$, where $s = \log x$. As $\overline{F} \in RV_{-1/\xi}$, the slope $\Psi'(s)$ converges to $1/\xi$ as $s \to \infty$. The log-log transform of the associated SV function L_F is given by $\Psi_{L_F}(s) = s/\xi - \Psi(s)$. Its slope Ψ'_{L_F} tends to 0 and measures the speed at which the influence of the slowly varying nuisance part vanishes.

Analogously, for the tail quantile function $U(t) = F^{\leftarrow}(1 - 1/t)$, we obtain $U(t) = e^{\Psi^{\leftarrow}(\log t)}$, where $\Psi^{\leftarrow}$ denotes the generalized inverse of Ψ (see Definition 3.1) and we write

$$U(t) = e^{\varphi(\log t)}, \quad \varphi(r) = \Psi^{\leftarrow}(r) = \int_1^{e^r} \frac{\varepsilon(y)}{y} dy + c, \tag{10}$$

with $\varepsilon(y) = yU'(y)/U(y)$ and $c = \log U(1)$.

As $\varphi(\Psi(r)) = r$, simple calculus shows that the sufficient first and second order conditions in the ultimate and penultimate approximation (see Theorems 3.1 and 3.2) may be equivalently expressed in terms of $\overline{F}$ or U.

Asymptotics for the g-and-h distribution

A random variable X is said to have a *g-and-h distribution*, if X satisfies

$$X = a + bk(Z) = a + b\frac{e^{gZ} - 1}{g} e^{hZ^2/2}, \quad a, g, h \in \mathbb{R}, b > 0,$$

where $Z \sim N(0, 1)$. We concentrate on the case $a = 0$ and $b = 1$. Degen et al. [10] show that for $g, h > 0$, the g-and-h distribution is regularly varying with index $-1/h$, i.e. $\overline{F}(x) = \overline{\Phi}(k^{-1}(x)) = x^{-1/h} L_F(x)$. Since F is differentiable, we may write

$$\overline{F}(x) = \overline{\Phi}(k^{-1}(x)) = e^{-\Psi(\log x)},$$

where, as above, Ψ denotes the log-log transform of $\overline{F}$.

By definition of k we have $y := \log k(s) = hs^2/2 + gs - \log g + O\left(e^{-gs}\right)$, $s \to \infty$. We rewrite this as $s^2 = 2y/h - 2gs/h + 2\log g/h + O\left(e^{-gs}\right)$, $s \to \infty$ and hence

$$s = k^{-1}(e^y) = \sqrt{\frac{2y}{h}} - \frac{g}{h} + \frac{1}{\sqrt{y}}\left(\frac{g^2}{(2h)^{3/2}} + \frac{\log g}{\sqrt{2h}}\right) + O\left(\frac{1}{y^{3/2}}\right), \quad y \to \infty.$$

Recall the standard asymptotic expansion for the normal distribution tail given by $\overline{\Phi}(x) = e^{-x^2/2}/(\sqrt{2\pi}x)\left(1 - 1/x^2 + O\left(1/x^4\right)\right)$, $x \to \infty$; see Abramowitz and Stegun [1], p. 932. Hence we obtain the following asymptotics for the g-and-h tail $\overline{F}$:

$$\begin{aligned} -\Psi(t) &= \log \overline{\Phi}(k^{-1}(e^t)) \\ &= -\frac{1}{2}\left(k^{-1}(e^t)\right)^2 - \log k^{-1}(e^t) - \log 2\pi + O\left(\frac{1}{(k^{-1}(e^t))^2}\right), \end{aligned}$$

as $t \to \infty$, which then leads to,

$$-\Psi(t) = -\frac{1}{h}t + \frac{\sqrt{2}g}{h^{3/2}}\sqrt{t} - \frac{1}{2}\log t - c + O\left(\frac{1}{\sqrt{t}}\right), \quad t \to \infty, \tag{11}$$

with $c = \frac{1}{2}\log\frac{4\pi}{h} + \frac{g^2}{h^2} + \frac{\log g}{h}$.

Asymptotics for the loggamma distribution

A random variable X follows a loggamma distribution with parameters α and β, if its density satisfies

$$f(x) = \frac{\alpha^\beta}{\Gamma(\beta)}(\log x)^{\beta-1}x^{-\alpha-1}, \quad \alpha, \beta > 0, x > 1,$$

where Γ denotes the Gamma function. The corresponding tail is given by

$$\overline{F}(x) = \frac{\Gamma(\beta, \alpha \log x)}{\Gamma(\beta)} = e^{-\Psi(\log x)}, \quad x > 1,$$

where

$$\Gamma(\beta, \alpha \log x) = \int_{\alpha \log x}^{\infty} t^{\beta-1}e^{-t}dt, \quad x > 1,$$

denotes the upper incomplete Gamma function. Therefore we obtain $-\Psi(t) = \log\Gamma(\beta, \alpha t) - \log\Gamma(\beta)$. For $\beta = 1$ we are in the Pareto case with $-\Psi(t) = -\alpha t$. For $\beta \in \mathbb{R}_+ \setminus \{1\}$, using the standard asymptotic expansion for the upper incomplete Gamma function given in Abramowitz and Stegun [1], p. 263, we obtain

$$-\Psi(t) = -\alpha t + (\beta - 1)\log t + c + O\left(\frac{1}{t}\right), \quad t \to \infty,$$

where $c = (\beta - 1)\log\alpha - \log\Gamma(\beta)$.