The Quantitative Modeling of Operational Risk: Between g-and-h and EVT

Matthias Degen[1] Paul Embrechts[2] Dominik D. Lambrigger[3]
Department of Mathematics, ETH Zurich, CH-8092 Zurich, Switzerland

First version: November 9, 2006
This version: December 19, 2006

Abstract

Operational risk has become an important risk component in the banking and insurance world. The availability of (few) reasonable data sets has given some authors the opportunity to analyze operational risk data and to propose different models for quantification. As proposed in Dutta and Perry [10], the parametric g-and-h distribution has recently emerged as an interesting candidate.

In our paper, we discuss some fundamental properties of the g-and-h distribution and their link to extreme value theory (EVT). We show that for the g-and-h distribution, convergence of the excess distribution to the generalized Pareto distribution (GPD) is extremely slow and therefore quantile estimation using EVT may lead to inaccurate results if data are well modeled by a g-and-h distribution. We further discuss the subadditivity property of Value-at-Risk (VaR) for g-and-h random variables and show that for reasonable g and h parameter values, superadditivity may appear when estimating high quantiles. Finally, we look at the g-and-h distribution in the one-claim-causes-ruin paradigm.

Keywords: Extreme Value Theory, g-and-h Distribution, Hill Estimator, LDA, Operational Risk, Peaks Over Threshold, Second Order Regular Variation, Subadditivity, Value-at-Risk.

[1] degen@math.ethz.ch
[2] embrechts@math.ethz.ch
[3] dominik.lambrigger@math.ethz.ch

1 Introduction

Since the early discussion around Basel II and Solvency II, the pros and cons of a quantitative (Pillar I) approach to operational risk have been widely put forward. Some papers, like Daníelsson et al. [6], have early on warned against an over optimistic view that tools from market (and to some extent credit) risk management can easily be transported to the Basel II framework for operational risk. Also, the actuarial community working on Solvency 2 so far defied a precise definition, and as a consequence a quantitative capital measurement for operational risk. The situation in the banking world is very different indeed, not only did Basel II settle on a precise definition, "The risk of loss resulting from inadequate or failed internal processes, people and systems or from external events. Including legal risk, but excluding strategic and reputational risk.", also concrete suggestions for risk capital calculation have been made. These include the basic-indicator (BI) approach, the standardized (S) approach and the loss distribution approach (LDA). BI and S are easy to calculate pure volume based measures. In the LDA however, banks are basically given full methodological freedom for the calculation of regulatory capital. The main reason being that for this new and especially from a statistical data point of view poorly understood risk class, regulators hoped that modeling freedom would yield a healthy competition among the quant groups of various financial institutions. Whereas this point of view is no doubt a laudable one, the imposed boundary conditions made a practical implementation more than difficult. Some of these constraints are the use of the risk measure (VaR), the level (99.9%) and the "holding" period (1 year). Of these, the extremely high quantile (corresponding to a 1 in 1000 year event estimation) is no doubt the most critical one. Beyond these, banks are required to augment internal data modeling with external data and expert opinion. The fact that current data—especially at the individual bank level—are far from being of high quality or abundant, makes a reliable LDA for the moment questionable.

By now, numerous papers, reports, software, textbooks have been written on the subject. For our purposes, as textbooks we would like to mention McNeil et al. [18] and Panjer [24]. Both books stress the relevance of the actuarial methodology towards a successful LDA; it is no coincidence that in McNeil et al. [18], Chapter 10 carries

the title "Operational Risk and Insurance Analytics". Another recent actuarial text that at some point will no doubt leave its footprint on the LDA platform is Bühlmann and Gisler [4].

For the present paper, two fundamental papers, which are center stage to the whole LDA controversy, are Moscadelli [22] and Dutta and Perry [10]. Both are very competently written papers championing different analytic approaches to the capital charge problem. Whereas Moscadelli [22] is strongly based on EVT, Dutta and Perry [10] introduce as a benchmark model the parametric g-and-h distribution. Moscadelli [22] concludes that, based on the 2002 Loss Data Collection Exercise (LDCE) of the Basel Committee, EVT yields reasonable capital estimates when data are pooled at Business Line (BL) level. A somewhat broader range for BL β-coefficients in the standardized approach beyond the (12-18)% range is suggested. The overall $\alpha = 15\%$ coefficient in the basic indicator approach is corroborated. The information coming through from individual banks with respect to the use of EVT is mixed. As explained in Nešlehová et al. [23], the statistical properties of the data are no doubt a main fact underlying this diffuse image. When it comes to high quantile estimation (and 99.9% is very high) EVT emerges as a very natural key methodological player; more on this later in the paper.

In Dutta and Perry [10] the authors conclude that "Many different techniques being tested by researchers are centered around EVT. In many of those cases we observe that attempts are made to fit a distribution or apply a method without understanding the characteristics of the loss data or the limitation of the models". And further, "In our experiment we found that EVT often did not work for data where we observed many large losses". Based on the 2004 LDCE, Dutta and Perry [10] suggest the g-and-h distribution as a viable option.

As already stated above, we consider both Moscadelli [22] and Dutta and Perry [10] as very well written. The latter paper also introduces a fundamental, more qualitative yardstick against which any capital charge model ought to be tested:

1. Good Fit - Statistically, how well does the method fit the data?

2. Realistic - If a method fits well in a statistical sense, does it generate a loss distribution with a realistic capital estimate?

3. Well Specified - Are the characteristics of the fitted data similar to the loss data and logically consistent?

4. Flexible - How well is the method able to reasonably accommodate a wide variety of empirical loss data?

5. Simple - Is the method easy to apply in practice, and is it easy to generate random numbers for the purposes of loss simulation?

In our paper, we will mainly look critically at the very recent g-and-h approach and compare and contrast its properties with EVT based methodology. As academics we do not possess real operational risk data so that our comments may be "academic" in nature; we do however hope that the various results discussed will contribute positively towards the quest for a reliable (in the sense of 1.-5. above) capital charge for operational risk. Based on the empirical findings of Dutta and Perry [10] that (1) operational risk data seem to be modeled appropriately by the g-and-h; and that (2) the EVT based Peaks Over Threshold (POT) approach does not seem to model the data well, we like to bridge these findings with theory.

We expect the reader to have studied Moscadelli [22] and Dutta and Perry [10] in detail. A basic textbook for EVT in the context of insurance and finance is Embrechts et al. [11]; see also Chapter 7 in McNeil et al. [18]. Before we start our discussion we find it worthwhile to put the record straight on EVT: Papers like Diebold et al. [8] and Dutta and Perry [10] highlight weaknesses of EVT when it comes to some real applications, especially in finance. In Embrechts et al. [11] these points were already stressed very explicitly. Like any statistical method, EVT (for instance in its Peaks Over Threshold (POT) or Hill estimator variant) only promises to deliver when a very precise set of conditions is satisfied. We strongly advice the reader to revisit Embrechts et al. [11] and look carefully at the following examples: Figure 4.1.13, Figure 5.5.4 and Figure 6.4.11. Nešlehová et al. [23] yields further warnings when EVT

is applied blindly to operational risk data. We very much hope that some of these underlying issues will become more clear when we progress through the paper.

The paper is organized as follows. In Section 2 we recall the definition of the g-and-h distribution and discuss some fundamental first order regular variation properties. In Section 3 we focus on second order regular variation as well as on the (slow) rate of convergence of a relevant subclass of the g-and-h excess distribution functions to the GPD. Further we analyze the impact of this on quantile estimation via the POT method. Subadditivity properties of VaR for g-and-h distributed random variables (rvs) are discussed in Section 4. In Section 5 we highlight the link between g-and-h and the one-claim-causes-ruin phenomenon. We conclude in Section 6.

2 The g-and-h distribution

2.1 The basic definition

Throughout this paper, rvs are denoted by capital letters $X_1, X_2, \ldots$ and assumed to be defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. These rvs will typically represent one-period risk factors in a quantitative risk management context. The next definition is basic to the analysis in Dutta and Perry [10].

DEFINITION 2.1 Let $Z \sim \mathcal{N}(0,1)$ be a standard normal rv. A rv X is said to have a g-and-h distribution with parameters $a, b, g, h \in \mathbb{R}$, if X satisfies

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

with the obvious interpretation for $g = 0$. We write $X \sim$ g-and-h, or when X has distribution function (df) F, $F \sim$ g-and-h. □

Instead of g and h being constants, a more flexible choice of parameters may be achieved by considering g and h to be polynomials including higher orders of Z^2. In Dutta and Perry [10], such a polynomial choice was necessary for some banks and business lines. For our paper, we restrict our attention to the basic case where g and h are constants. The parameters g and h govern the skewness and the heavy-tailedness of the distribution, respectively; see Hoaglin et al. [14].

In the case $h = 0$, equation (1) reduces to $X = a + b\dfrac{e^{gZ} - 1}{g}$, which is referred to as the g-distribution. The g-distribution thus corresponds to a scaled lognormal distribution. In the case $g = 0$, equation (1) is interpreted as $X = a + bZe^{hZ^2/2}$, which is referred to as the h-distribution. The case $g = h = 0$ corresponds to the normal case. The linear transformation parameters a and b are of minor importance for our purposes. Unless otherwise stated we restrict our attention to the g-and-h distribution with parameters $a = 0$ and $b = 1$. Furthermore we assume $g, h > 0$. Parameters of the g-and-h distributions used in Dutta and Perry [10] to model operational risk (at enterprise level) are within the following ranges: $g \in [1.79, 2.30]$ and $h \in [0.10, 0.35]$.

REMARK 2.1 Since the function $k(x) = \frac{e^{gx}-1}{g}e^{hx^2/2}$ for $h > 0$ is strictly increasing, the df F of a g-and-h rv X can be written as

$$F(x) = \Phi(k^{-1}(x)),$$

where Φ denotes the standard normal df. This representation immediately yields an easy procedure to calculate quantiles and hence the Value-at-Risk of a g-and-h rv X,

$$\mathrm{VaR}_\alpha(X) = F^{-1}(\alpha) = k(\Phi^{-1}(\alpha)), \quad 0 < \alpha < 1. \qquad \square$$

In the next section we derive some properties of the g-and-h distribution which are important for understanding its estimation properties of high quantiles.

2.2 Tail properties and regular variation

In questions on high quantile estimation, the statistical properties of the estimators used very much depend on the tail behavior of the underlying model. The g-and-h distribution is very flexible in that respect. There are numerous graphical techniques for revealing tail behavior of dfs. We restrict our attention to mean excess plots (me-plots) and log-log density plots. In Figure 1 we show a me-plot for a g-and-h distribution with parameter values typical in the context of operational risk. Besides the thick line corresponding to the theoretical mean excess function, we plot 12 empirical mean excess functions based on $n = 10^5$ simulated g-and-h data. The upward sloping behavior of the me-plots indicates heavy-tailedness as typically present in the

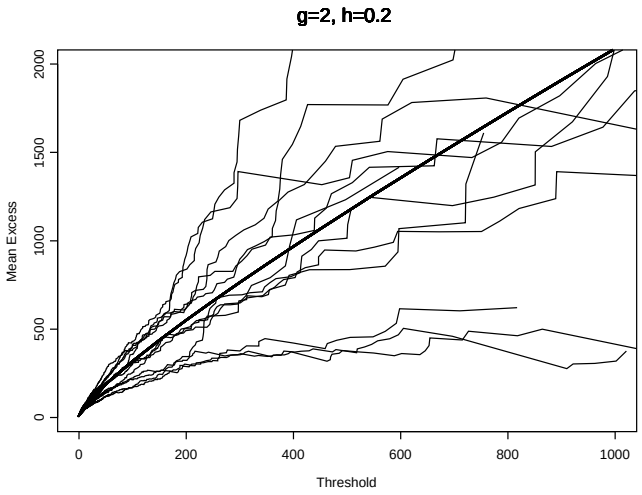

Figure 1: Theoretical mean excess function (thick line) together with 12 empirical mean excess plots of the g-and-h distribution.

class of subexponential dfs $\mathcal{S}$ (see Embrechts et al. [11], Figure 6.2.4), linear behavior corresponding to Pareto (power) tails. In the latter case, the resulting log-log-density plot shows a downward sloping linear behavior; see Figure 2 for a typical example. Figure 1 also highlights the well-known problem when interpreting me-plots, i.e. a very high variability of the extreme observations made visible through the simulated me-plots from the same underlying model. Both figures give insight into the asymptotic heavy-tailedness of the g-and-h. We now make this property analytically precise.

A standard theory for describing heavy-tailed behavior of statistical models is Karamata's theory of regular variation. For a detailed treatment of the theory, see Bingham et al. [5]. Embrechts et al. [11] contains a summary useful for our purposes. Recall that a measurable function $L : \mathbb{R} \to (0, \infty)$ is slowly varying (denoted $L \in SV$) if for $t > 0$:

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

A function f is called regularly varying (at ∞) with index $\alpha \in \mathbb{R}$ if $f(x) = x^\alpha L(x)$ and is denoted by $f \in RV_\alpha$; note that $RV_0 = SV$. The following proposition is an immediate consequence of Karamata's Theorem; see Embrechts et al. [11], Theorem A3.6. It provides an easy tool for checking regular variation. In the context of EVT,

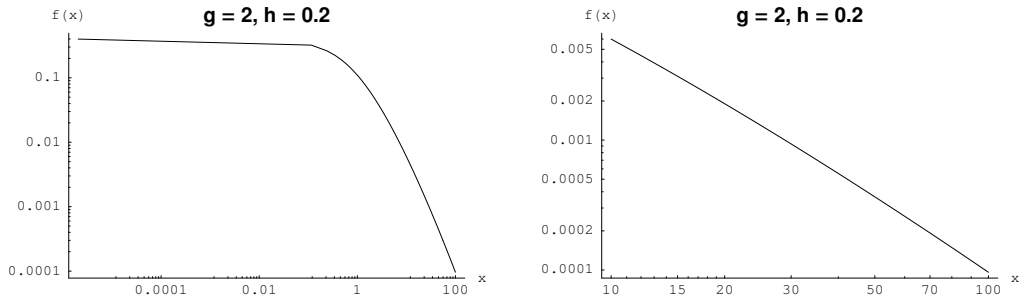

Figure 2: Density of the g-and-h distribution plotted on a log-log scale. Note the different plotting ranges of the axes.

the result is known as *von Mises condition* for the Fréchet df; see Embrechts et al. [11], Corollary 3.3.8. Throughout we denote $\overline{F} = 1 - F$.

Proposition 2.1 *Let F be an absolutely continuous df with density f satisfying*

$$\lim_{x\to\infty} \frac{xf(x)}{\overline{F}(x)} = \alpha > 0,$$

then $\overline{F} \in RV_{-\alpha}$. □

Note that we slightly abuse the notation, in the sense that we should restrict RV to non-negative rvs. Through tail-equivalence (see Embrechts et al. [11], Definition 3.3.3) we can easily bypass this issue.

We proceed by showing that the g-and-h distribution is indeed regularly varying (at ∞) with index $-1/h$ (still assuming $h > 0$). Assume $X \sim$ g-and-h, then

$$\begin{aligned} F(x) &= \Phi(k^{-1}(x)), \\ f(x) &= \frac{\varphi(k^{-1}(x))}{k'(k^{-1}(x))}, \end{aligned}$$

where φ denotes the density of a standard normal rv. Using Mill's ratio $u(1-\Phi(u))/\varphi(u) \to 1$, as $u \to \infty$, we have that

$$\begin{aligned}
\lim_{x\to\infty} \frac{xf(x)}{\overline{F}(x)} &= \lim_{x\to\infty} \frac{x\varphi(k^{-1}(x))}{(1-\Phi(k^{-1}(x)))k'(k^{-1}(x))} \\
&\overset{u:=k^{-1}(x)}{=} \lim_{u\to\infty} \frac{k(u)\varphi(u)}{(1-\Phi(u))k'(u)} \\
&= \lim_{u\to\infty} \frac{\varphi(u)(e^{gu}-1)}{(1-\Phi(u))(ge^{gu}+hu(e^{gu}-1))} \\
&= \frac{1}{h}
\end{aligned}$$

and hence $\overline{F} \in RV_{-1/h}$.

REMARK 2.2 In a similar way, one shows that also the h-distribution ($h > 0$) is regularly varying with the same index. This was already mentioned in Morgenthaler and Tukey [21]. The g-distribution ($g > 0$) however is—as a scaled lognormal distribution—subexponential but not regularly varying. At this point the reader is advised to have a look at Section 1.3.2 and Appendix A 3.2 in Embrechts et al. [11], or Section 5 later in the paper. □

In summary, we have the following result.

Theorem 2.1

Suppose $F \sim$ g-and-h with $g, h > 0$, then $\overline{F} \in RV_{-1/h}$. For $h = 0$ and $g > 0$, we have $F \in \mathcal{S}\backslash RV$, where $\mathcal{S}$ denotes the class of subexponential dfs. □

Hence, if $X \sim$ g-and-h with $h > 0$ we have by definition of regular variation $\overline{F}(x) = x^{-1/h}L(x)$ for some slowly varying function L. A key message from our paper is that the precise behavior of L may profoundly affect the statistical properties of EVT-based high quantile estimations. This point was very clearly stressed in Embrechts et al. [11]; see Figure 4.1.13 and Example 4.1.12. End-users in risk management and financial applications seem to have missed out on this message. We will show how absolutely crucial this point is. The quality of high quantile estimation for power tail data very much depends on the second order behavior of the underlying (often unknown) slowly varying function L.

Below we derive an explicit asymptotic formula for the slowly varying function L in the case of the g-and-h distribution. For $g, h > 0$ we have

$$\begin{aligned} L(x) &= \overline{F}(x)x^{1/h} \\ &= (1 - \Phi(k^{-1}(x)))x^{1/h}. \end{aligned}$$

and hence

$$\begin{aligned} L(k(x)) &= (1 - \Phi(x))(k(x))^{1/h} \\ &= (1 - \Phi(x)) \left(\frac{e^{gx} - 1}{g} \right)^{1/h} e^{x^2/2} \\ &= \frac{1}{\sqrt{2\pi}x} \left(\frac{e^{gx} - 1}{g} \right)^{1/h} \left(1 + O\left(\frac{1}{x^2} \right) \right), \end{aligned}$$

leading to

$$L(x) = \frac{1}{\sqrt{2\pi}g^{1/h}} \frac{\left(e^{gk^{-1}(x)} - 1 \right)^{1/h}}{k^{-1}(x)} \left(1 + O\left(\frac{1}{(k^{-1}(x))^2} \right) \right), \quad x \to \infty.$$

In order to find an explicit asymptotic inverse function of k, define

$$\tilde{k}(x) = \frac{1}{g} e^{h\frac{x^2}{2} + gx} \sim k(x), \quad x \to \infty,$$

with inverse function

$$\tilde{k}^{-1}(x) = -\frac{g}{h} + \frac{1}{h}\sqrt{g^2 + 2h\log(gx)}, \quad x > 0.$$

Here and throughout the paper, $f(x) \sim g(x)$, $x \to a$ means that $\lim_{x \to a} \frac{f(x)}{g(x)} = 1$. Note that $\tilde{k}^{-1}(x) \sim k^{-1}(x)$ for $x \to \infty$. Altogether we obtain:

Theorem 2.2

Let F be the df of a g-and-h rv with $g, h > 0$. Then $\overline{F}(x) = x^{-1/h}L(x)$, where for $x \to \infty$,

$$L(x) = \frac{1}{\sqrt{2\pi}g^{1/h}} \frac{\left[\exp\left(g\left(-\frac{g}{h} + \frac{1}{h}\sqrt{g^2 + 2h\log(gx)} \right) \right) - 1 \right]^{1/h}}{-\frac{g}{h} + \frac{1}{h}\sqrt{g^2 + 2h\log(gx)}} \left(1 + O\left(\frac{1}{\log x} \right) \right).$$

PROOF: Define

$$\tilde{L}(x) = \frac{1}{\sqrt{2\pi}g^{1/h}} \frac{\left(e^{g\tilde{k}^{-1}(x)} - 1\right)^{1/h}}{\tilde{k}^{-1}(x)}.$$

Note that $u := k^{-1}(x)$ is a strictly increasing function for $g, h > 0$. Hence,

$$\begin{aligned}
\frac{L(x)}{\tilde{L}(x)} &= \frac{\sqrt{2\pi}g^{1/h}\tilde{k}^{-1}(x)(1-\Phi(k^{-1}(x)))x^{1/h}}{\left(e^{g\tilde{k}^{-1}(x)} - 1\right)^{1/h}} \\
&= \frac{\sqrt{2\pi}g^{1/h}\tilde{k}^{-1}(k(u))(1-\Phi(u))(k(u))^{1/h}}{\left(e^{g\tilde{k}^{-1}(k(u))} - 1\right)^{1/h}} \\
&= \left(\frac{e^{gu}-1}{e^{g\tilde{k}^{-1}(k(u))}-1}\right)^{1/h} \frac{\tilde{k}^{-1}(k(u))(1-\Phi(u))}{\varphi(u)} \\
&= 1 + O\left(\frac{1}{u^2}\right) \\
&= 1 + O\left(\frac{1}{\log x}\right), \quad x \to \infty
\end{aligned}$$

which completes the proof. □

REMARK 2.3 The slowly varying function L in the above theorem is (modulo constants) essentially of the form $\exp(\sqrt{\log x})/\sqrt{\log x}$. This will turn out to be a particularly difficult type of slowly varying function in the context of EVT. □

In the next section we will look more carefully into this issue by studying the second order behavior of L. For this we will first make a link to EVT and discuss how the properties of L may influence the statistical estimation of high quantiles based on EVT.

3 Second order regular variation

3.1 The Pickands-Balkema-de Haan Theorem

We assume the reader to be familiar with univariate EVT. The notation used in this section is adopted from Embrechts et al. [11]. For a g-and-h rv X (with $g, h > 0$) it was shown in the previous section that $F \in \mathrm{MDA}(H_\xi)$, i.e. belongs to the maximum domain of attraction of an extreme value distribution with index $\xi = h > 0$. The

Pickands-Balkema-de Haan Theorem, Theorem 3.4.13(b) in Embrechts et al. [11], implies that for $F \in \text{MDA}(H_\xi)$ there exists a positive measurable function $\beta(\cdot)$, such that

$$\lim_{u \uparrow x_0} \sup_{x \in (0, x_0 - u)} \left| F_u(x) - G_{\xi, \beta(u)}(x) \right| = 0.$$

We denote the upper support point of F by x_0. In the case of a g-and-h distribution, $x_0 = \infty$. By the above theorem, the excess df F_u, defined by $F_u(x) = P(X - u \leq x | X > u)$, is well approximated by the df of a GPD, $G_{\xi,\beta(u)}$, for high threshold values u. This first-order convergence result stands at the heart of EVT. For practical purposes however second-order properties of F are of even greater importance (e.g. performance of parameter estimates or calculation of high quantiles). We are interested in the rate of convergence of F_u towards $G_{\xi,\beta(u)}$, i.e. in how fast does

$$d(u) := \sup_{x \in (0, x_0 - u)} \left| F_u(x) - G_{\xi, \beta(u)}(x) \right|$$

converge to 0 for $u \to x_0$. Define

$$\begin{aligned} V(t) &:= (1 - F)^{-1}(e^{-t}) \\ A(t) &:= \frac{V''(\log t)}{V'(\log t)} - \xi, \end{aligned}$$

for some $F \in \text{MDA}(H_\xi)$. The following proposition (see Raoult and Worms [25], Corollary 1) gives insight into the behavior of the growth rate $d(u)$ in cases including, for example, the g-and-h distribution with $\xi = h > 0$.

Proposition 3.1 *Let $F \in MDA(H_\xi)$ be a df which is twice differentiable and let $\xi > -1$. If the following conditions are satisfied:*

i) $\lim_{t \to \infty} A(t) = 0$,

ii) $A(\cdot)$ *is of constant sign near* ∞,

iii) there exists $\rho \leq 0$ *such that* $|A| \in \text{RV}_\rho$,

then, for $u \to x_0$,

$$d(u) := \sup_{x \in (0, x_0 - u)} \left| F_u(x) - G_{\xi, V'(V^{-1}(u))}(x) \right| = O(A(e^{V^{-1}(u)})).$$

□

The parameter ρ is called the second order regular variation parameter. Recall that for a g-and-h distribution $F(x) = \Phi(k^{-1}(x))$ and hence $\overline{F}^{-1}(x) = k(\Phi^{-1}(1-x))$. In this case the function V defined above is given by $V(t) = k(\Phi^{-1}(1-e^{-t}))$.
Moreover,

$$V'(\log t) = \frac{k'(\nu(t))}{t\varphi(\nu(t))}$$

and

$$V''(\log t) = \frac{k''(\nu(t)) - tk'(\nu(t))\Big(\varphi(\nu(t)) + \varphi'(\nu(t))/(t\varphi(\nu(t)))\Big)}{(t\varphi(\nu(t))^2},$$

where $\nu(t) := \Phi^{-1}(1-\frac{1}{t})$. One easily checks conditions i) and ii) above. In addition, using Lemma 2 of Raoult and Worms [25], it can be shown that $|A| \in RV_\rho$ with second order parameter $\rho = 0$. By definition of V we have

$$\begin{aligned} A\big(e^{V^{-1}(u)}\big) &= \frac{V''(\log e^{V^{-1}(u)})}{V'(\log e^{V^{-1}(u)})} - h = \frac{V''(\log 1/\overline{F}(u))}{V'(\log 1/\overline{F}(u))} - h \\ &= \frac{k''(k^{-1}(u))\overline{F}(u)}{k'(k^{-1}(u))\varphi(k^{-1}(u))} + \frac{k^{-1}(u)\overline{F}(u)}{\varphi(k^{-1}(u))} - 1 - h. \end{aligned}$$

Lemma 3.1 *For $X \sim$ g-and-h with $g, h > 0$, the following asymptotic relation holds:*

$$A\Big(e^{V^{-1}(k(x))}\Big) \sim \frac{g}{x}, \quad x \to \infty.$$

PROOF: Using Mill's ratio, we have

$$\begin{aligned} \lim_{x\to\infty} \frac{A\big(e^{V^{-1}(k(x))}\big)}{g/x} &= \frac{1}{g}\lim_{x\to\infty}\left(\frac{k''(x)}{k'(x)}\frac{x\overline{\Phi}(x)}{\varphi(x)} + x\left(\frac{x\overline{\Phi}(x)}{\varphi(x)} - 1\right) - hx\right) \\ &= \frac{1}{g}\lim_{x\to\infty}\left(\frac{k''(x)}{k'(x)}\left(1 + O\left(\frac{1}{x^2}\right)\right) - hx\right) \\ &= \frac{1}{g}\lim_{x\to\infty}\left(\frac{h^2x^2/g + 2hx + g + h/g}{hx/g + 1}\left(1 + O\left(\frac{1}{x^2}\right)\right) - hx\right) \\ &= \frac{1}{g}\lim_{x\to\infty}\frac{h + O(1/x)}{h/g + 1/x} = 1. \end{aligned}$$

□

By Proposition 3.1 and since $k^{-1}(\cdot)$ is increasing (still assuming $g, h > 0$), the rate of convergence of the excess df of a g-and-h distributed rv towards the GPD $G_{\xi,\beta(u)}$ with $\xi = h$ and $\beta(u) = V'(V^{-1}(u))$ is given by

$$d(u) = O\Big(\frac{1}{k^{-1}(u)}\Big) = O\Big(\frac{1}{\sqrt{\log u}}\Big), \quad u \to \infty.$$

At this point we would like to stress that $d(u) = O(\frac{1}{\sqrt{\log u}})$ does not imply that the rate of convergence is independent of the parameters g and h. Not a detailed derivation of this fact, but rather a heuristic argument is provided by the following:

$$\frac{\log L(x)}{\log x} \sim \sqrt{2}\frac{g}{h^{3/2}}\frac{1}{\sqrt{\log x}} = O(\frac{1}{\sqrt{\log x}}), \quad x \to \infty.$$

Clearly the value $\frac{g}{h^{3/2}}$ affects the rate of convergence of $\log L(x)/\log x$ as $x \to \infty$.

In Table 1 we have summarized the rates of convergence in the GPD approximation as a function of the underlying df. For both the exponential as well as the exact Pareto, $d(u) = 0$. For dfs like the double exponential parent, normal, Student t and Weibull, convergence is at a reasonably fast rate. Already for the very popular lognormal and loggamma dfs, convergence is very slow. This situation deteriorates further for the g-and-h where the convergence is extremely slow. Note that one can always construct dfs with arbitrary slow convergence of the excess df towards the GPD; see Resnick [26], Exercise 2.4.7. This result is in a violent contrast to the rate of convergence in the Central Limit Theorem which, for finite variance rvs, is always $n^{-1/2}$.

Distribution	Parameters	$\overline{F}(x)$	ρ	$d(u)$
Exponential(λ)	$\lambda > 0$	$e^{-\lambda x}$	$-\infty$	0
Pareto(α)	$\alpha > 0$	$x^{-\alpha}$	$-\infty$	0
Double exp. parent		e^{-e^x}	-1	$O(e^{-u})$
Student t	$\nu > 0$	$\bar{t}_\nu(x)$[1]	$-2/\nu$	$O(\frac{1}{u^2})$
Normal(0, 1)		$\overline{\Phi}(x)$	0	$O(\frac{1}{u^2})$
Weibull(τ, c)	$\tau \in \mathbb{R}_+ \setminus \{1\}, c > 0$	$e^{-(cx)^\tau}$	0	$O(u^{-\tau})$
Lognormal(μ, σ)	$\mu \in \mathbb{R}, \sigma > 0$	$\overline{\Phi}(\frac{\log x - \mu}{\sigma})$	0	$O(\frac{1}{\log u})$
Loggamma(γ, α)	$\alpha > 0, \gamma \neq 1$	$\Gamma(\gamma, \alpha \log x)$	0	$O(\frac{1}{\log u})$
g-and-h	g,h > 0	$\overline{\Phi}(k^{-1}(x))$	0	$O(\frac{1}{\sqrt{\log u}})$

Table 1: Rate of convergence to the GPD for different distributions, as a function of the threshold u.

[1] $\bar{t}_\nu(x) \sim c(\nu)x\left(1 + x^2/\nu\right)^{-(\nu+1)/2}$

From a theoretical point of view this already yields a first important result: if data are well modeled by a g-and-h distribution with $g, h > 0$, then high quantile estimation for such data based on the POT method will typically converge very slowly. In the next section we will look at this issue in somewhat more detail.

It is often stated by some authors that they have solved the (critical) optimal choice of threshold problem in the POT or Hill method. On several occasions we have stressed that this problem has *no* general solution; optimality can only be obtained under some precise second order properties on the underlying slowly varying function L (we concentrate on the Fréchet case). It is precisely this L (let alone its second order properties) which is impossible to infer from statistical data. Hence, the choice of a reasonable threshold (we avoid using the word "optimal") remains the Achilles heel of any high quantile estimation procedure based on EVT. For a more pedagogic and entertaining presentation of the underlying issues, see Embrechts and Nešlehová [12].

3.2 Threshold choice

There exists a huge literature on the optimal threshold selection problem in EVT; see for instance Beirlant et al. [3] for a review. Within a capital charge calculation problem, the choice of threshold u above which EVT fits well the tail of the underlying df *may* significantly influence the value estimated. We stress the word "may"; indeed in some cases the quantile estimate is rather insensitive with respect to the choice of u, in other cases it is very sensitive. This stresses the fact that for the modeling of extremes, great care as to the underlying model and data properties has to be taken. The analysis below is indicative of the underlying issues and definitely warrants a much broader discussion. We have included it to warn the reader for some of the difficulties in using automatic procedures for determining so-called "optimal" tail regions for the estimation of high quantiles. We restrict our attention to g-and-h dfs and estimate quantiles using the Hill estimator. The conclusions obtained also hold for the MLE based POT method.

We assume that $X_1, X_2, \ldots, X_n$ are iid realizations from a continuous df F with $\overline{F} \in RV_{-1/\xi}$, i.e. $\overline{F}(x) = x^{-1/\xi} L(x)$, $L \in SV$.

DEFINITION 3.1 The *Hill estimator* is defined by

$$H_{k,n} := \frac{1}{k}\sum_{j=1}^{k} \log X_{n-j+1,n} - \log X_{n-k,n} \quad (1 < k < n),$$

where $X_{1,n} \leq \ldots \leq X_{n,n}$ are the order statistics of $X_1, \ldots, X_n$. □

Define the quantile function $U(x) := F^{-1}(1 - 1/x)$. Since $\overline{F} \in RV_{-1/\xi}$, we have $U(x) = x^{\xi} l(x)$, for some slowly varying function l; see for instance Beirlant et al. [3]. If there exist $\rho \leq 0$ and a positive function b with $b(x) \to 0$ for $x \to \infty$, such that for all $t \geq 0$

$$\log \frac{l(tx)}{l(x)} \sim b(x) k_\rho(t), \quad x \to \infty,$$

with

$$k_\rho(t) \;=\; \begin{cases} \frac{t^\rho - 1}{\rho} & \rho < 0, \\ \log t & \rho = 0, \end{cases}$$

then the asymptotic mean square error (AMSE) of the Hill estimator satisfies

$$\begin{aligned} \mathrm{AMSE} H_{k,n} &:= (\mathrm{ABias} H_{k,n})^2 + \mathrm{AVar} H_{k,n} \\ &= \left(\frac{b\big((n+1)/(k+1)\big)}{1-\rho} \right)^2 + \frac{\xi^2}{k}; \end{aligned} \tag{2}$$

see for instance Matthys and Beirlant [16]. Applying this result to the regularly varying g-and-h df F with index $1/\xi = 1/h \in (0, \infty)$, we get

$$\begin{aligned} l(x) &= \frac{1}{g(2\pi)^{h/2}} \frac{e^{g\Phi^{-1}(1-1/x)} - 1}{(\Phi^{-1}(1-1/x))^h} \\ &= \frac{1}{g(2\pi)^{h/2}} \frac{e^{g\Phi^{-1}(1-1/x)}}{(\Phi^{-1}(1-1/x))^h} (1 - e^{-g\Phi^{-1}(1-1/x)}). \end{aligned}$$

Using the following approximation for the quantile function of the normal,

$$\Phi^{-1}\Big(1 - \frac{1}{x}\Big) \sim \sqrt{\log \frac{x^2}{2\pi} - \log\log \frac{x^2}{2\pi}}, \quad x \to \infty,$$

(see e.g. Dominici [9], Proposition 21) we arrive at

$$\begin{aligned} \log \frac{l(tx)}{l(x)} &= g\left(\Phi^{-1}(1-1/(tx)) - \Phi^{-1}(1-1/x)\right) \\ &\quad + h \log \frac{\Phi^{-1}(1-1/x)}{\Phi^{-1}(1-1/(tx))} + \log \frac{1 - e^{-g\Phi^{-1}(1-1/(tx))}}{1 - e^{-g\Phi^{-1}(1-1/x)}} \\ &= \left[\frac{g}{(2\log x)^{1/2}} - \frac{h}{2\log x} - \frac{g}{(2\log x)^{3/2}} + o\left(\frac{1}{(\log x)^{3/2}}\right) \right] \log t, \quad x \to \infty. \end{aligned}$$

Hence, in particular, $\rho = 0$ for the g-and-h distribution with $g, h > 0$. Note that, given the precise model assumptions above, formula (2) yields an obvious approach to estimating the optimal sample fraction to calculate the Hill estimator.

In practice however, one usually does not have any information about the second order properties of the underlying df. Thus for each k, $1 < k < n$, $b(\cdot), \rho$ and h have to be estimated form the data by estimators $\widehat{b}(\cdot)$, $\widehat{\rho}$ and $\widehat{h}$, which are for example the maximum likelihood or the least squares estimators; see Beirlant et al. [3]. One then chooses k in the following way:

$$k_{opt} = \underset{k \in \mathbb{N}_+}{\operatorname{argmin}} \left[\left(\frac{\widehat{b}\big((n+1)/(k+1)\big)}{1-\widehat{\rho}} \right)^2 + \frac{\widehat{h}^2}{k} \right].$$

We will now apply this procedure to simulated g-and-h data. For each pair of parameter values g and h (see Table 2 below) we simulate a hundred samples of 2000 observations from a g-and-h distribution. For each of the 100 samples we compute the Hill estimator $\widehat{h}^{Hill}_{k_{opt}}$ of h using k_{opt} number of upper order statistics of the 2000 observations of that sample. For every cell in Table 2 we thus get 100 estimates $\left(\widehat{h}^{Hill}_{k_{opt},m}\right)_{1 \le m \le 100}$ of h. To analyze the performance of the Hill estimator $\widehat{h}^{Hill}_{k_{opt}}$ we calculate the standardized root mean square error (SRMSE), which for a single cell is given by

$$\frac{1}{h} \sqrt{\frac{1}{100} \sum_{m=1}^{100} \left(\widehat{h}^{Hill}_{k_{opt},m} - h \right)^2}.$$

The SRMSE of the Hill estimator $\widehat{h}^{Hill}_{k_{opt}}$ is summarized in Table 2. From there we may deduce a characteristic pattern which essentially remains the same for other threshold selection procedures. We confirmed this by implementing the optimal threshold selection method proposed by Guillou and Hall [13] and by further applying an ad-hoc selection method, using a fixed percentage of exceedances of 5%.

An identical study was performed for the MLE estimates $\widehat{h}^{MLE}$ of h, yielding very similar results to the case of $\widehat{h}^{Hill}_{k_{opt}}$. Therefore, whether using Hill or MLE to estimate h, the key message we infer from Table 2 is that EVT-based tail index estimation leads to highly inaccurate results. Moreover, the larger the ratio g/h, the larger the SRMSE. In particular, for parameter values reported in Dutta and Perry [10], which are in a range around $g = 2, h = 0.2$, the SRMSE is close to 400%. The

$g\backslash h$	0.1	0.2	0.5	0.7	1	2
0.1	142	82	33	23	18	11
0.2	165	97	42	32	25	20
0.5	224	132	49	38	27	19
0.7	307	170	63	44	29	20
1	369	218	86	58	36	26
2	696	385	151	108	74	31
3	1097	613	243	163	115	54

Table 2: SRMSE (in %) of the Hill estimator $\widehat{h}^{Hill}_{k_{opt}}$ of h for g-and-h data for different parameter values of g and h.

numbers reported in Table 2 are somewhat counterintuitive. Indeed in papers like McNeil and Saladin [19] and Dutta and Perry [10] it is stated that heavier tailed models require higher thresholds and likewise a larger sample size to achieve a similar error bound. Table 2 on the other hand indicates, that for fixed g the SRMSE decreases for increasingly heavier tails.

The poor performance of EVT-based tail index estimation, especially for parameter values with a large ratio g/h, e.g. $g = 2$ and $h = 0.2$, is further confirmed by a Hill plot; see Figure 3 (right panel). On the other hand, we expect a "good" Hill plot for g/h small, e.g. $g = 0.1$ and $h = 1$, which is confirmed by Figure 3 (left panel). In the left panel, the Hill plot is rather flat over a large range of threshold values yielding an accurate estimate of the true value $h = 1$. In the right panel however, the Hill plot is absolutely misleading. Though being temptingly flat, an estimation of the shape parameter $\xi = h$ based on the Hill plot would in that case lead to a completely wrong estimate of $\widehat{h}^{Hill} \approx 0.7$, whereas the true value is $h = 0.2$. One can easily come up with finite mean g-and-h examples (i.e. $h < 1$) leading to infinite mean EVT estimates ($\widehat{h}^{Hill} > 1$). Such an example can be constructed by choosing the skewness parameter sufficiently high. We exemplify this issue in Figure 4, where we present a Hill plot for $n = 10^6$ realizations of a g-and-h rv with parameter values $g = 4$ and $h = 0.2$ (finite mean). Again the Hill plot shows a relatively flat behavior, suggesting a value of $\widehat{h}^{Hill} \approx 1.2$, indicating infinite mean.

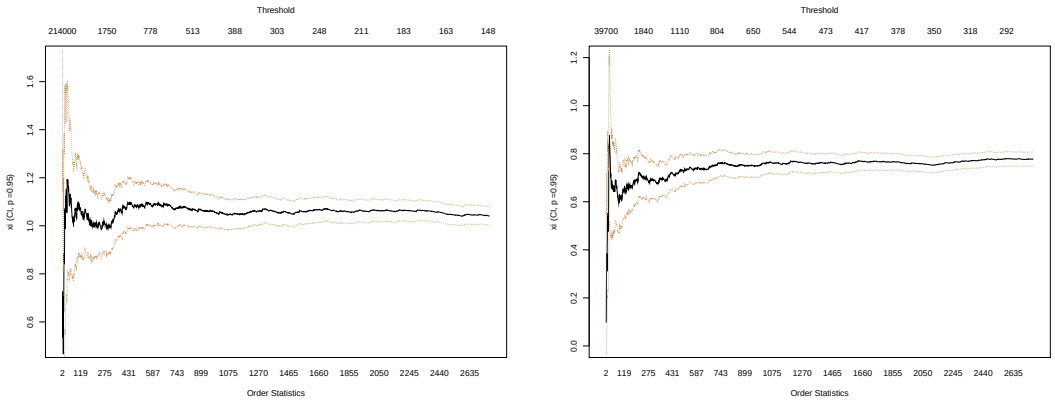

Figure 3: Hill plot for $g = 0.1, h = 1$ and $n = 10^6$ (left panel) and $g = 2, h = 0.2$ and $n = 10^6$ (right panel).

In summary, given that data are well modeled by a g-and-h distribution where g/h is large, as is the case with the data reported by Dutta and Perry [10], an EVT based estimation of the tail index h unavoidably leads to highly inaccurate estimates. Consequently in such cases high quantile estimation using standard EVT methodology becomes highly sensitive to specific numerical estimation procedures. We emphasize this further in the next section.

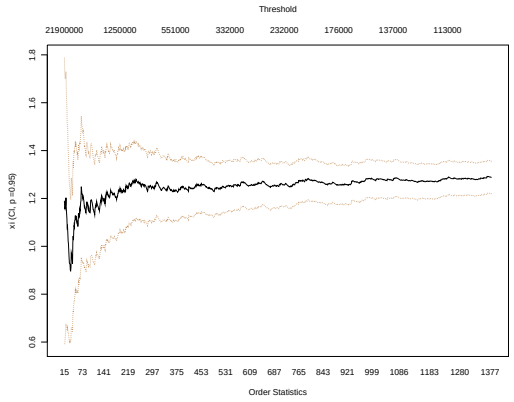

Figure 4: Hill plot for $g = 4, h = 0.2$ and $n = 10^6$.

3.3 Quantile estimation

To confirm our findings of the previous section we performed a quantile estimation study along the lines of McNeil and Saladin [19], [20]. Instead of applying sophisticated optimal threshold selection procedures we likewise concentrated on an ad-hoc method by taking into account only a certain percentage of the highest data points; see McNeil and Saladin [19] for details. We generated g-and-h data by rejection sampling and calculated the POT estimator of the 99% and the 99.9% quantiles for different values of g and h. We compared our results ($h = 0.2$ and $h = 1$) with the findings of McNeil and Saladin [19] to conclude that the performance of the POT estimator for the g-and-h distribution is much worse—in terms of high standardized bias and SRMSE—than for any of the distributions used in that paper.

From a methodological point of view, Makarov [15] is also relevant in this respect. In that paper, the author shows that uniform relative quantile convergence in the Pickands-Balkema-de Haan Theorem necessarily needs a slowly varying function L which is asymptotically constant. Clearly, L in the g-and-h case is far from being constant.

All the results shown so far point to the fact that the slowly varying function L for the g-and-h distribution for $g, h > 0$ renders high quantile estimation based on EVT methodology difficult: for g-and-h type data, all EVT based procedures show extremely slow convergence and hence these estimators may be highly inaccurate.

In order to better understand the relative merits of EVT and g-and-h, we now turn to estimating quantiles in cases where EVT is known to do well and see how g-and-h based estimation compares. In the Tables 3 and 4 we give the estimated quantiles for two empirical data sets; the daily S&P data from 1960 to 1993 and the Danish fire insurance data from 1980 to 1990, as discussed in Embrechts et al. [11]. We compare empirical quantile estimation, POT quantile estimation and the g-and-h method. For the latter we fit a g-and-h distribution to the data, where we allow for location and scale parameters to be different from $a = 0$, $b = 1$. The parameters a, b, g, h are estimated using Tukey's percentiles. Using the language of Hoaglin et al. [14], we take approximately $\log_2(n)$ letter values, where n is the number of available data points, with the full spread (FS) for the S&P data and with the upper half

spread (UHS) for the Danish fire insurance data; see for instance Hoaglin et al. [14] and Dutta and Perry [10], Appendix C. The quantile is then given by $a + bk(\Phi^{-1}(\alpha))$.

	Empirical	POT	g-and-h
90%	0.93	1.10	0.92
95%	1.30	1.34	1.29
99%	2.14	2.13	2.23
99.9%	4.10	4.30	3.98

Table 3: Quantile estimation of S&P-data with $n = 8414$ data points. In the case of the POT-Method we fix the threshold to $u = 1.45$.

	Empirical	POT	g-and-h
90%	5.54	5.64	5.72
95%	9.97	9.30	9.43
99%	26.04	27.51	27.32
99.9%	131.55	121.17	101.51

Table 4: Quantile estimation of Danish fire insurance data with $n = 2167$ data points. In the case of the POT-Method we fix the threshold to $u = 5$.

We conclude that for the 95% and 99% levels all methods yield rather similar results, wheras for very high quantiles, the results differ substantially. Of course for the S&P data a more dynamic modeling, as for instance done in McNeil and Frey [17] including careful backtesting, would be useful. In the case of the Danish data backtesting to find the better fitting procedure is not really available. Once more, these results are in no way conclusive. We have included them to highlight some issues and hopefully encourage further research.

As a final comparison we test the three quantile estimation methods mentioned above by means of two selected examples in line with McNeil et al. [18], Section 7.2.5. We will distinguish between a "soft" and a "hard" problem. With regards to the "soft"

problem, we generate 1000 realizations of a standard normal rv and estimate the 95%-quantile, whereas for the "hard" problem we generate 1000 realizations of a t_3 rv and estimate the 99.9%-quantile. So in the "soft" problem we estimate a quantile well within the range of light tailed data. For the "hard" problem we estimate a quantile at the edge of heavy tailed data. In both problems our estimations are based on the empirical, the POT and the g-and-h method by means of the procedure mentioned above. In the case of the g-and-h method the full spread is used to estimate the parameters a, b, g, h. In Figure 5 we plot the SRMSE as a function of the chosen threshold of the GPD.

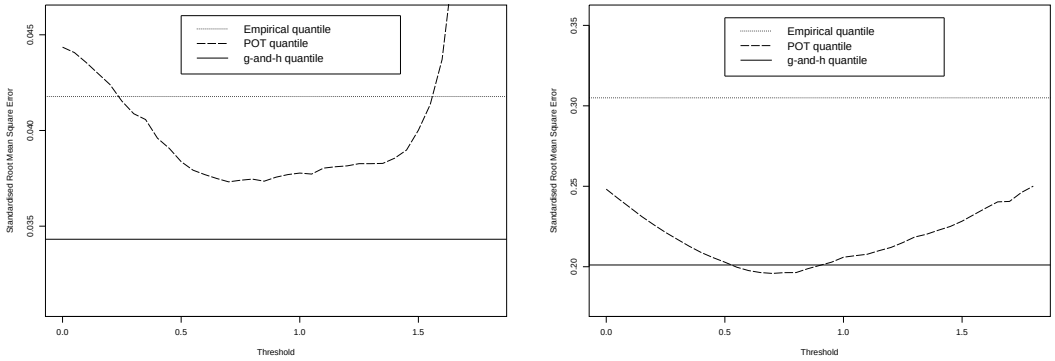

Figure 5: SRMSE for the "soft" case (left panel) and for the "hard" case (right panel).

In the soft case, where the quantile is estimated at a moderate level, g-and-h fits well and its SRMSE is smaller than the SRMSE obtained by the POT method. This is not surprising, as the normal distribution perfectly fits into the g-and-h framework. In the hard case the g-and-h method as well as the POT method clearly outperforms the empirical estimator.

4 Subadditivity of VaR

As stated above, we can give an explicit formula for the Value-at-Risk in the case of a g-and-h rv:

$$\mathrm{VaR}_\alpha(X) = k(\Phi^{-1}(\alpha)), \quad 0 < \alpha < 1,$$

with

$$k(x) = \frac{e^{gx} - 1}{g} e^{hx^2/2}.$$

In Dutta and Perry [10] the authors state: "We have not mathematically verified the subadditivity property for g-and-h, but in all cases we have observed empirically that enterprise level capital is less than or equal to the sum of the capitals from business lines or event types". Of course, a mathematical discussion of subadditivity would involve multivariate modeling; we will return to this issue in a future publication.

In order to statistically investigate the subadditivity property for the g-and-h distribution, we perform a simulation study. Let X_1, X_2 be iid g-and-h rvs with parameters $g = 2.4$ and $h = 0.2$. We estimate (by simulation of $n = 10^7$ realizations) the diversification benefit $\delta_{g,h}(\alpha) = \mathrm{VaR}_\alpha(X_1) + \mathrm{VaR}_\alpha(X_2) - \mathrm{VaR}_\alpha(X_1 + X_2)$, where of course $\delta_{g,h}(\alpha)$ will be non-negative if and only if subadditivity occurs. Our results are displayed in Figure 6. For the above realistic choice of parameters superadditivity holds for α smaller than a certain level $\widetilde{\alpha} \approx 99.4\%$. The fact that subadditivity, i.e. $\mathrm{VaR}_\alpha(X_1 + X_2) \leq \mathrm{VaR}_\alpha(X_1) + \mathrm{VaR}_\alpha(X_2)$, holds for α sufficiently large is well known; see Proposition 4.1 below. That superadditivity enters for typical operational risk parameters at levels below some $\widetilde{\alpha}$ may be somewhat surprising. The latter may be important in the discussion around the scaling of risk measures. Indeed, risk managers realize that estimating VaR_α at level $\alpha \geq 99\%$, say, is statistically difficult. It has been suggested to estimate VaR_α deeper down in the data, $\alpha = 90\%$, say, and then scale up to 99.9%. The change from super- to subadditivity over this range should be of concern.

Note that one can even construct finite-mean examples (choosing the skewness parameter g large enough) for levels $\widetilde{\alpha} = 99.9\%$ and higher, such that subadditivity of Value-at-Risk fails for all $\alpha < \widetilde{\alpha}$. This should be viewed in contrast to the following proposition by Daníelsson et al. [7]. See also that paper for a definition of bivariate

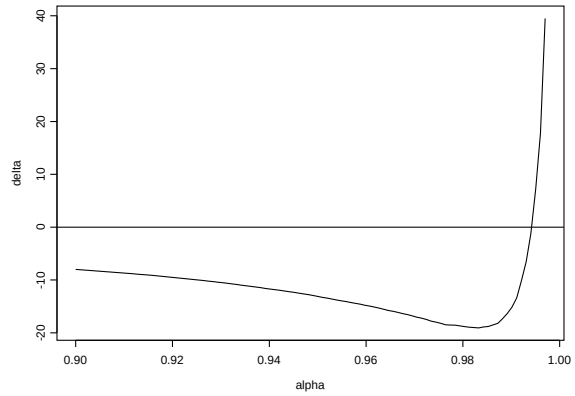

Figure 6: Plot of $\delta_{g,h}(\alpha)$ as a function of α for $g = 2.4, h = 0.2; n = 10^7$.

regular variation.

Proposition 4.1 *Suppose that the non-degenerate vector* (X_1, X_2) *is regularly varying with extreme value index* $\xi < 1$. *Then* VaR_α *is subadditive for* α *sufficiently large.* □

Figure 6 exemplifies the subadditivity of VaR *only* in the very upper tail region. The reader should thus be warned that Proposition 4.1 is an asymptotic statement and does not guarantee subadditivity for a broad range of high quantiles. Furthermore, note that for $\xi = h > 1$ subadditivity typically fails. The reason being that for $h > 1$ one deals with infinite mean models; see Nešlehová et al. [23] for more details on this.

For practitioners it will be of prime importance to know for which choices of g and h values one can expect subadditivity. As shown in Figure 6, this depends on the level α. We restrict ourself to the α-values 99% and 99.9%, relevant for practice. Assume that the operational risk data of two business lines of a bank are well modeled by iid g-and-h rvs with parameter values $g \in [1.85, 2.30]$, $h \in [0.15, 0.35]$. Note that these values roughly correspond to the parameters estimated by Dutta and Perry [10] at *enterprise* level. It would be of interest to figure out if aggregation at business line

level leads to diversification in the sense of subadditivity of VaR. For this purpose we consider two iid g-and-h rvs with g and h values within the afore mentioned ranges. In Figure 7 we display a contour plot of $\delta_{g,h}(\alpha)$ for a fixed α, together with the rectangle containing the parameter values of interest.

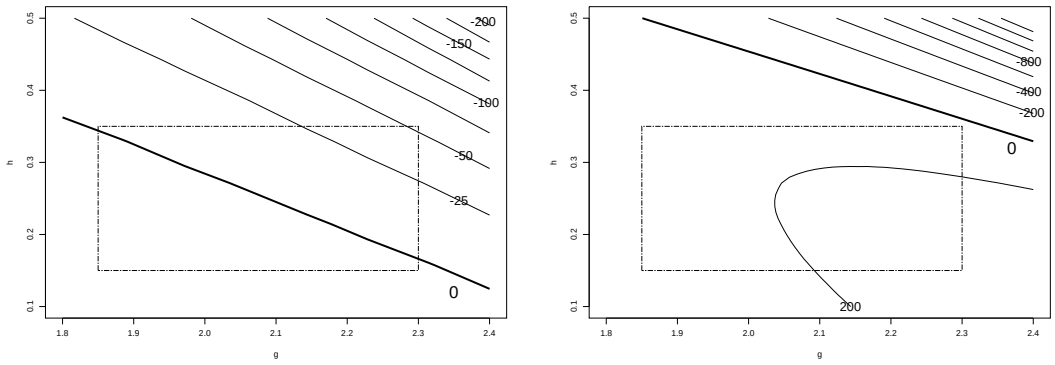

Figure 7: Contour plot of $\delta_{g,h}(\alpha)$ as a function of g and h for fixed $\alpha = 99\%$ (left panel) and $\alpha = 99.9\%$ (right panel); $n = 10^7$.

The number attached to each contour line gives the value of $\delta_{g,h}(\alpha)$ and the lines indicate levels of equal magnitude of diversification benefit. The 0-value corresponds to these models where VaR_α is additive, $\mathrm{VaR}_\alpha(X_1 + X_2) = \mathrm{VaR}_\alpha(X_1) + \mathrm{VaR}_\alpha(X_2)$. The positive values (bottom left hand corner) correspond to models yielding subadditivity. The top right hand corner, corresponding to negative values for $\delta_{g,h}(\alpha)$, leads to superadditivity for the corresponding parameter values. Note that for $\alpha = 99.9\%$, the entire parameter rectangle lies within the region of subadditivity; see right panel of Figure 7. It is though important to realize that with only relatively small changes in the underlying g and h parameters, one may end up in the superadditivity region. The situation becomes more dramatic at lower quantiles. The left panel of Figure 7 corresponds to $\alpha = 99\%$ (which is still relatively high!). There the superadditivity region extends and a substantial fraction of our parameter rectangle lies therein.

The above statements were made under the iid assumption. In the example below we allow for dependence. For this we link the marginal g-and-h distributions with the

same parameters as in Figure 6 by a Gauss-copula; see McNeil et al. [18], p. 191. In Figure 8 we plot $\delta_{g,h}(\alpha)$ for three different correlation parameters $\rho = 0, 0.5$ and 0.7. This figure should be compared with Figure 6.

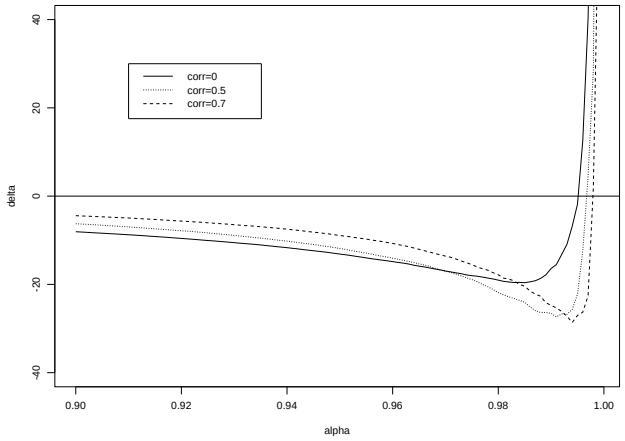

Figure 8: Plot of $\delta_{g,h}(\alpha)$ as a function of α for $g = 2.4, h = 0.2, n = 10^7$ and Gauss-copula with correlation $\rho = 0, 0.5$ and 0.7. Note that $\rho = 0$ corresponds to the independence case in Figure 6.

It appears that in a range below 95%, $|\delta_{g,h}(\alpha)|$ becomes smaller when the correlation parameter increases. This is not surprising because VaR is additive under comonotonic dependence, i.e. for risks with maximal correlation; see McNeil et al. [18], Theorem 5.25. As a consequence $\delta_{g,h}(\alpha)$ would be tending to 0 for $\rho \to 1$. The effect of dependence can clearly be seen for large values of α. Based on our simulation study, it appears that with increasing correlation ρ, the range of superadditivity extends to even higher values of α. Hence the stronger the dependence the higher the level α has to be in order to achieve a subadditive model. Formulated differently, for strong dependence (ρ large), most levels α chosen in practice will lie within the range of superadditivity. We have worked out these results also for other dependence structures, like the t- and the Gumbel-copula. For these cases we also elaborated contour plots as in Figure 7. The results do not differ significantly from Figure 7 and thus we refrain from displaying these plots here.

The situation in any LDA is of course in general much more complicated than in our simple example above. Practitioners and risk managers should therefore interpret our statements rather from a methodological and pedagogical point of view. It seems that diversification of operational risk can go the wrong way due to the skewness and heavy-tailedness of this type of data.

5 The one-claim-causes-ruin paradigm

In several publications on operational risk it is stated that often relatively few claims cause the major part of the total operational risk loss. Papers highlighting this phenomenon in an operational risk context are Nešlehová et al. [23] and Böcker and Klüppelberg [2]. Though these publications contain the relevant results, for matter of completeness we reformulate the main conclusions in terms of the g-and-h distribution. We concentrate on the iid case, changes incorporating dependence between the different loss rvs along the lines of Böcker and Klüppelberg [2] can easily be made.

Let $X_1, \ldots, X_d$ be iid g-and-h rvs and $S_d = \sum_{i=1}^{d} X_i$ the total loss. Recall that for $g > 0, h \geq 0$ the g-and-h distribution is subexponential, i.e.

$$\mathbb{P}[S_d > x] \sim \mathbb{P}[\max_{1 \leq i \leq d} X_i > x], \quad x \to \infty.$$

The above relation expresses the fact that for subexponential distributions the tail distribution of the total loss S_d is determined by the tail distribution of the maximum loss. We are in the so-called "one-claim-causes-ruin" regime; see Embrechts et al. [11] or Asmussen [1].

More generally, consider $(X_i)_{i \geq 0}$ a sequence of iid g-and-h rvs, independent of a counting process $(N_t)_{t \geq 0}$ and $S_t = \sum_{i=1}^{N_t} X_i$. Hence we have

$$G_t(x) := \mathbb{P}[S_t \leq x] = \sum_{n=0}^{\infty} \mathbb{P}[N_t = n] F^{n*}(x),$$

where F^{n*} denotes the n-th convolution of F. Furthermore, by Theorem 1.3.9 of Embrechts et al. [11], if there exists $\varepsilon > 0$ such that

$$\sum_{n=0}^{\infty} (1+\varepsilon)^n \mathbb{P}[N_t = n] < \infty, \tag{3}$$

then the tail df of S_t satisfies

$$\mathbb{P}[S_t > x] \sim \mathbb{E}[N_t]\overline{F}(x), \quad x \to \infty.$$

Note that condition (3) is for instance satisfied in the Poisson, Binomial and Negative Binomial case. The above representation implies

$$G_t^{-1}(\alpha) \sim \overline{F}^{-1}\left(\frac{1-\alpha}{\mathbb{E}[N_t]}\right), \quad \alpha \to 1,$$

and hence for $F \sim$ g-and-h with $g > 0, h \geq 0$,

$$\mathrm{VaR}_\alpha(S_t) \sim k\left(\Phi^{-1}\left(1 - \frac{1-\alpha}{\mathbb{E}[N_t]}\right)\right), \quad \alpha \to 1.$$

Though these results yield explicit analytic approximations for VaR_α, α large, their practical importance is questionable.

6 Conclusion

In this paper we have highlighted some of the modeling issues for g-and-h severity distributions within an LDA for operational risk. There seems to be a discrepancy in practice between results which strongly favor EVT methodology (Moscadelli [22]) and g-and-h methodology (Dutta and Perry [10]). Our main results are as follows. First, the g-and-h class of dfs yields an overall very slow rate of convergence in applications using EVT based techniques. This is mainly due to the second order behavior of the slowly varying function underlying the g-and-h for $h > 0$. As a consequence, setting an optimal threshold for an EVT based POT approach becomes very difficult and hence quantile (risk capital) estimates may become unreliable. Second, the issue of sub- or superadditivity of g-and-h based VaR estimation very much depends on the parameter values g and h. It is shown that, both for iid as well as for dependent data, small changes in the underlying parameters may lead VaR to switch regime (super to sub or vice versa). Finally, since the class of g-and-h distributions is subexponential (for $g > 0, h \geq 0$), this class of dfs also yields the one-claim-causes-ruin phenomenon.

Several of the above results (observations) were based on simulation studies. We do however believe that the message delivered in our paper may already have considerable relevance for practical application of the LDA for operational risk. In future

publications we shall come back to some of these issues in a more analytic form. In particular, we are working on QRM relevant properties of multivariate g-and-h models.

Acknowledgments

The authors would thank Valérie Chavez-Demoulin, Kabir Dutta, Johanna Nešlehová and Jason Perry for several discussions related to the quantitative modeling of operational risk and for the numerous useful comments which led to a substantial improvement of the paper.

References

[1] Asmussen, S. (2001) *Ruin Probabilities*. World Scientific, Singapore.

[2] Böcker, K. and Klüppelberg, C. (2006) Multivariate models for operational risk. Preprint, TU Munich.

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

[4] Bühlmann, H. and Gisler, A. (2005) *A Course in Credibility Theory and its Applications*. Springer, Berlin.

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

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

[7] Daníelsson, J., Jorgensen, B. N., Samorodnitsky, G., Sarma, M. and de Vries, C. G. (2005) Subadditivity re-examined: the case for Value-at-Risk. Preprint, London School of Economics.

[8] Diebold, F., Schuermann, T. and Stronghair, J. (2001) Pitfalls and opportunities in the use of extreme value theory in risk management. In: Refenes, A.-P., Moody, J. and Burgess, A. (Eds.), *Advances in Computational Finance*, Kluwer Academic Press, Amsterdam, pp. 3-12, Reprinted from: Journal of Risk Finance, 1, 30-36 (Winter 2000).

[9] Dominici, D. (2003) The inverse of the cumulative standard normal probability function. *Integral Transforms and Special Functions* **14**, 281–292.

[10] 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.

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

[12] Embrechts, P. and Nešlehová, J. (2006) Extreme value theory. Copulas. Two talks on the DVD *Quantitative Financial Risk Management. Fundamentals, Models and Techniques*, Henry Stewart.

[13] Guillou, A. and Hall, P. (2001) A diagnostic for selecting the threshold in extreme value analysis. *J.R. Statist. Soc. B* **63**, 293–305.

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

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

[16] Matthys, G. and Beirlant, J. (2000) Adaptive threshold selection in tail index estimation. In: Embrechts, P. (Ed.), *Extremes and Integrated Risk Management*, pp. 37–48, Risk Books, London.

[17] McNeil, A. J. and Frey, R. (2000) Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance* **7**, 271–300.

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

[19] McNeil, A. J. and Saladin, T. (1997) The peaks over thresholds method for estimating high quantiles of loss distributions. *Proceedings of 28th International ASTIN Colloquium.*

[20] McNeil, A. J. and Saladin, T. (2000) Developing scenarios for future extreme losses using the POT method. In: Embrechts, P. (Ed.), *Extremes and Integrated Risk Management*, pp. 253–267, Risk Books, London.

[21] Morgenthaler, S. and Tukey, J. W. (2000) Fitting quantiles: doubling, HR, HH and HHH distributions. *Journal of Computational and Graphical Statistics* **9**, 180–195.

[22] 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.

[23] 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.

[24] Panjer, H. H. (2006) *Operational Risk: Modeling Analytics*. Wiley, New York.

[25] Raoult, J.-P. and Worms, R. (2003) Rate of convergence for the generalized Pareto approximation of the excesses. *Advances of Applied Probability* **35**, 1007–1027.

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

Paul Embrechts
Department of Mathematics
ETH Zurich
CH-8092 Zurich, Switzerland
E-Mail: embrechts@math.ethz.ch