Infinite Mean Models and the LDA for Operational Risk

Johanna Nešlehová

RiskLab, Department of Mathematics
ETH Zurich
Rämistrasse 101
CH-8092 Zurich, Switzerland
Tel.: +41 44 632 3227
e-mail: johanna@math.ethz.ch
URL: http://www.math.ethz.ch/~johanna/

Paul Embrechts

Department of Mathematics
ETH Zurich
Rämistrasse 101
CH-8092 Zurich, Switzerland
Tel.: +41 44 632 3419
e-mail: embrechs@math.ethz.ch
URL: http://www.math.ethz.ch/~embrechts/

Valérie Chavez-Demoulin

Department of Mathematics
ETH Zurich
Rämistrasse 101
CH-8092 Zurich, Switzerland
Tel.: +41 44 632 3227
e-mail: valerie@math.ethz.ch
URL: http://www.math.ethz.ch/~valerie/

Abstract

Due to published statistical analyses of operational risk data, methodological approaches to the AMA modeling of operational risk can be discussed more in detail. In this paper we raise some issues concerning correlation (or diversification) effects, the use of extreme value theory and the overall quantitative risk management consequences of extremely heavy-tailed data. We especially highlight issues around infinite mean models. Besides methodological examples and simulation studies, the paper contains indications for further research.

Keywords: AMA, coherence, diversification, extremes, infinite mean models, LDA, operational risk, Pareto, subadditivity.

1 Introduction

The Advanced Measurement Approach (AMA) to the Pillar I modeling of Operational Risk, as defined in the Basel II proposals, raises some fundamental methodological issues. The common denominator of our comments concerns models for extremes and correlation. We do not strive for a full review of the underlying problems; several other publications have contributed to this. We rather make the operational risk modeler aware of issues which need a careful discussion. The reason for writing this paper grew out of numerous discussions we had with practitioners and academics alike, and this mainly because of the increasing availability of operational risk data. The topics presented are very much driven by our own research agenda; current developments on the operational risk scene have however accentuated their importance.

Recently, a practitioner made a comment to us that "It seems that the use of Extreme Value Theory (EVT) leads to counterintuitive results concerning the operational risk capital charge in the presence of diversification." Mainly due to analyses like Moscadelli (2004), de Fontnouvelle et al. (2004) and de Fontnouvelle (2005), considerable insight in the more detailed properties of operational risk data has been gained. One important consequence of these quantitative impact studies is the current interest in very heavy-tailed loss distributions. As EVT is increasingly been used by the insurance and finance industry in the realm of Quantitative Risk Management (QRM), it is more than ever necessary to recall some of the fundamental issues underlying any EVT analysis; this we do in Section 2, mainly based on several simulation examples. Section 3 recalls some examples touched upon in earlier work, especially in Embrechts et al. (2002), and stresses consequences for the correlation discussion and risk capital reduction in the light of portfolio diversification. Complementary to our discussions on EVT in Section 2 and correlation and diversification in Section 3, in Section 4 we recall some of the results from actuarial risk theory centered around the so-called "one claim causes ruin" phenomenon. Section 5 summarizes our findings and yields guidance for further research.

After the euphoria of a quantitative modeling (VaR-based) approach to market and credit risk, it was expected by some that a similar success could be achieved for operational risk. Whereas this eventually may prove to be true, at the moment there are serious arguments which need to be settled first. We want to bring some of these arguments into the open and avoid that the industry awakens to a statement like Mr Darcy in Jane Austen's Pride and Prejudice: "I see your design, Bingley, you dislike an argument, and want to silence this."

Besides EVT, research related to coherent risk measures and the modeling of dependence beyond correlation will be relevant for our discussion. The following publications play an essential role in the current debate on possibilities and limitations of QRM. First of all, Embrechts et al. (1997) early on promoted EVT as a useful set of techniques for the quantitative risk manager. Secondly, there is the work by Artzner et al. (1999) on the axiomatic theory of risk measures, and finally Embrechts et al. (2002) wrote a well received paper on properties and pitfalls of the use of correlation as a measure of dependence. The latter paper, first available as a RiskLab, ETH Zurich report in early 1999, was highly influential in starting risk managers to think beyond linear correlation, and stood at the cradle of the current copula-revival. A textbook reference where all these issues are summarized, put into a historical perspective and brought to bear on QRM is McNeil et al. (2005); the latter publication also contains an extensive bibliography from which hints for further reading can be obtained.

2 EVT and high-quantile estimation

In its most stylized form, the AMA for operational risk concerns the calculation of a risk measure ρ_γ at a confidence level γ for a loss random variable (rv) L. The latter corresponds to the aggregate losses over a given period Δ. Under the Basel II guidelines for operational risk, $\gamma = 99.9\%$, $\rho_\gamma = \mathrm{VaR}_\gamma$, the Value-at-Risk at level γ, and Δ equals one year. The loss rv typically has the form

$$L = \sum_{k=1}^{d} L_k \tag{1}$$

where $L_1, \ldots, L_d$ correspond to the loss rvs for given business lines and/or risk types as defined in the Basel II Accord. For the purpose of this section, the precise stochastic structure of the L_k's is less important; see however Section 4. It turns out to be essential that, as reported in detail in Moscadelli (2004) and to some extent in de Fontnouvelle et al. (2004), the tails of the loss distribution functions (dfs) are in first approximation heavy-tailed Pareto-type, i.e.

$$\mathsf{P}(L_k > x) = x^{-\alpha_k} h_k(x), \quad k = 1, \ldots, d, \tag{2}$$

where $\alpha_k > 0$ is the *tail-index* parameter and h_k is a *slowly varying* function; see Embrechts et al. (1997), p. 564. For an intuitive understanding of the meaning of (2), it helps to know that for $\varepsilon > 0$, $\mathrm{E}(L_k^{\alpha_k-\varepsilon}) < \infty$ and $\mathrm{E}(L_k^{\alpha_k+\varepsilon}) = \infty$, so that whenever $0 < \alpha_k < 1$, $\mathrm{E}(L_k) = \infty$ and for $1 < \alpha_k < 2$, $\mathrm{E}(L_k) < \infty$ but $\mathrm{var}(L_k) = \infty$. When analysed across the $d = 8$ business lines, the quantitative impact study data, as reported by Moscadelli (2004), bolster up (2) with 6 out of 8 business lines yielding estimates $0 < \widehat{\alpha}_k < 1$, i.e. *infinite mean models*. For three business lines, the results are statistically significant at the 95% level: corporate finance ($1/\widehat{\alpha} = 1.19$), trading & sales ($1/\widehat{\alpha} = 1.17$) and payment & settlement ($1/\widehat{\alpha} = 1.23$). These findings are somewhat put in perspective in de Fontnouvelle et al. (2004) where it is argued that at the level of individual banks, a slightly less (but still) heavy-tailed regime results. An operational risk capital charge is then based on the calculation of $\mathrm{VaR}_\gamma(L)$ with $\gamma = 99.9\%$.

A crucial problem in need of a solution is therefore a high-quantile ($\mathrm{VaR}_{99.9\%}$, say) estimation for heavy-tailed data; EVT provides a natural statistical toolkit for tackling this problem. It is not our aim to recall the methodology underlying the EVT-based approach to high-quantile estimation, we rather expect the reader to be familiar with the basics. Numerous monographs and papers have been written on the application of EVT to problems in financial and insurance risk management; we refer the interested reader to Embrechts et al. (1997), McNeil et al. (2005), Chavez-Demoulin et al. (2006) and the references therein for details and further reading. In what follows, we rather give examples on some issues regarding extremely heavy tails which recently came up in discussions on applications of EVT to operational risk modeling. The examples given provide as much an agenda for future research in EVT as some practical guidelines/warnings for the application of EVT methodology to operational risk data.

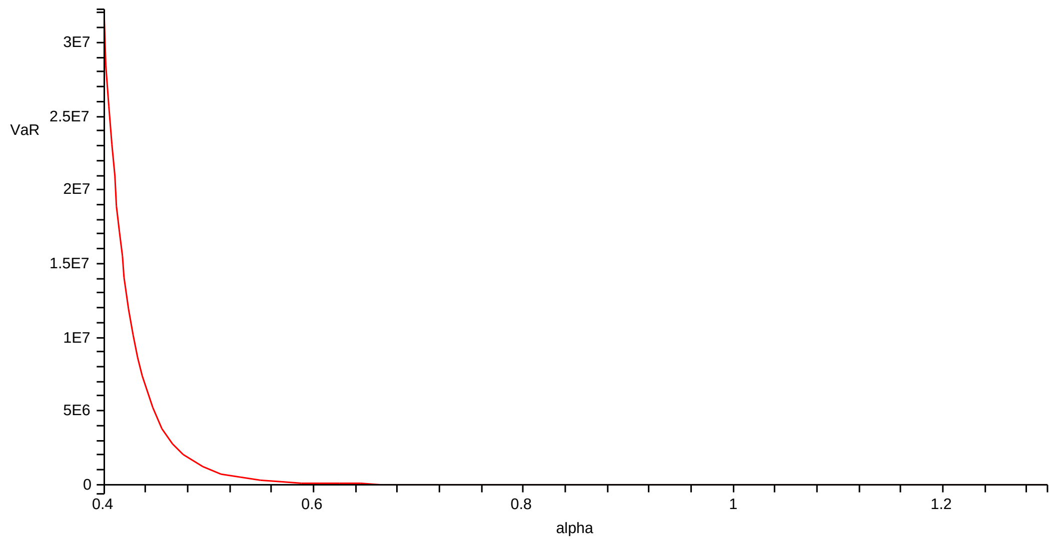

Figure 1: Value-at-Risk as a function of the tail-index parameter for an exact Pareto loss random variable.

A consequence of our findings will be that EVT should not be used blindly. As we will see in the next section, infinite mean models cause serious problems regarding diversification. A first issue we want to bring to the risk modeler's attention is that Value-at-Risk is sensitive to the values of α, especially for α small. In Figure 1, we plotted $\mathrm{VaR}_{99.9\%}$ of an exact Pareto loss rv L, i.e. $\mathsf{P}(L > x) = x^{-\alpha}$, $x \geq 1$, as a function of the tail-index parameter α. In this case, $\mathrm{VaR}_{99.9\%}$ is given by $0.001^{-1/\alpha}$, i.e. the Value-at-Risk grows with the tail-index parameter at an *exponential* rate. A similar result holds if the exact Pareto assumption is replaced by (2), say. This means that VaR may lead to ridiculously high capital charges in the infinite mean case ($\alpha < 1$). One could even argue that infinite mean models ought be banned from the operational risk modelers toolkit! The previous sentence we deliberately wrote in a rather provocative way. It is not sufficient to say that "Of course such models cannot occur since total capital existing is finite." The financial and economic literature is full of power-law models; see for instance Mandelbrot (1997), Mandelbrot and Hudson (2004) and Rachev and Mittnik (2000). All power-law (i.e. Pareto-type) models have some infinite moments, and yet they are used frequently. The transition from infinite second and finite first moment, say, to infinite first moment (the mean) is however a very serious one and should be handled very carefully. As a lighthouse warning against dangerous shoals, careful statistical analysis combined with judgement can help in pointing at the possibility of such a transition in the data. Theoretically, VaR_γ can be calculated in the infinite mean case, but "stop and pause" when this happens, reflect on what is going on and definitely do not continue with a push-of-the-button risk management attitude.

At this point, we would like to stress once again that as soon as questions are addressed that concern regions far in the tail of the loss distribution, EVT inevitably has to enter the modeling in one way or the other. In the words of Richard Smith, "What EVT is doing is making the best use of whatever data you have about extreme phenomena." Furthermore, quoting from Coles (2001), "Notwithstanding objections to the general principle of extrapolation, there are no serious competitor models to those provided by extreme value theory." Consequently, no other methodology which would yield an alternative and suitable framework for high quantile estimation exists. Statistical inference for EVT models and the model choice itself is another matter, however. As we will see with the examples below, an inappropriate statistical analysis can indeed have quite misleading outcomes. In the light of what has been said above, especially situations where the tail-index parameter estimates take values in $(0, 1)$ at least call for judgment. Therefore, we first take a step back and consider some possible reasons for such results.

- *The loss distribution is indeed extremely heavy-tailed, i.e. with infinite mean.* We believe that in this situation it is more than ever true that mathematics is only one part of the puzzle. A good understanding of the data generating process is crucial, and we can only voice a *warning* against any method that by means of magical or mathematical tricks manicures the tail behavior of the loss distribution. Below, we present some examples, which we believe may prove useful in the quest for a better understanding of such data.

- *Incorrect statistical inference.* Mistakes of this kind may be due to several reasons. It has been pointed out by numerous authors, that the Achilles heel of almost any EVT-based data analysis is the choice of the "right" threshold, which is both not too low to cause significant bias and not too high to lead to significant variance. It is not our aim to discuss this complex issue in further detail. We merely illustrate in Example 2.1 below, how a bad threshold choice can lead to under-/overestimation of VaR. A further issue is for instance the choice of the estimator of the tail-index parameter α. In Figure 2 we

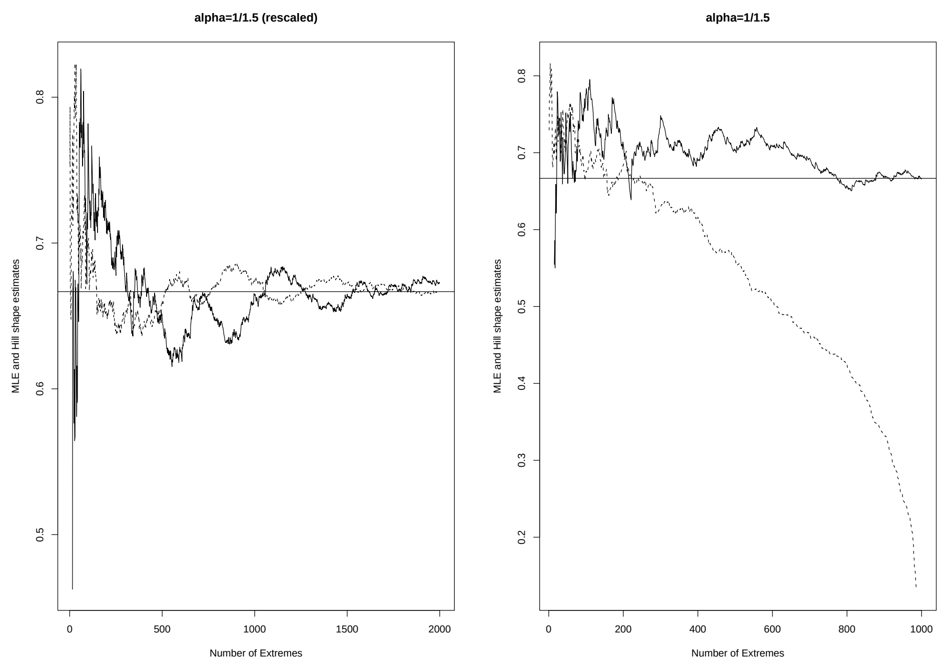

Figure 2: Hill (dotted line) vs. MLE (full line) estimators for iid data following the exact Pareto distribution with $\alpha = 2/3$ (left plot) and the generalized Pareto distribution with $\alpha = 2/3$ and $\beta = 1$ (right plot).

compare the behavior of the Hill and MLE(-GPD) estimators. Even for well-behaved data sets for which both estimators can be justified asymptotically, the difference can be quite substantial (if one uses the same threshold). Compounded with the findings in Figure 1, this can lead to considerable differences in reported risk capital.

- *EVT assumptions are not satisfied.* The basic EVT models require (close to) iid data, an assumption which is often too restrictive for practical applications. Therefore, extensions have been developed that allow for non-stationarity and/or clustering effects. Note that the use of such extensions in operational risk typically requires knowledge not only about the loss amount, but also about the exact loss occurence time. The latter may however not always be available in practice.

We now turn to some of the estimation issues in the for risk management “infinite mean” danger zone, viewed from an EVT perspective. The examples provided should be regarded as statistical warning lights when dealing with extremely heavy-tailed data.

In our opinion, a major cause of concern is *data contamination*! By the latter we mean that there are observations within the sample, which do not follow the same distribution as the rest of the data. Situations of this type have been considered in several mathematical fields: for instance *cluster analysis* and *robust statistics* to mention the most relevant ones. We do not intend to discuss either in further detail here; we refer the interested reader for instance to Hampel et al. (1986), Huber (1981) and McLachlan and Basford (1988). As the word "contamination" appears in numerous (mathematical as well as other) contexts in the literature, we would like to spare the reader any confusion by stressing that in our case, the use of this word does not go beyond the above meaning. Besides, the terminology is not meant to indicate that the contaminated data are a mere nuisance which should be prevented from taking influence upon the risk capital charge. We simply point out that special attention is called for when contamination is suspected. In the examples below, we consider two tailor-made models which we find relevant for operational risk. For instance, the event type which incorporates law suits may typically contain the largest losses. If all event types were thrown together, contamination as discussed in our first example would be expected in this case.

Example 2.1. Contamination above a very high threshold. First, we discuss a situation where the k largest values in the sample follow a different distribution as the remaining data. More specifically, we consider an iid sample $X_1, \ldots, X_n$ with order statistics $X_{1,n} \geq \cdots \geq X_{k,n} \geq \cdots \geq X_{n,n}$ drawn from a distribution of the form

$$F_X(x) = \begin{cases} 1 - \left(1 + \frac{x}{\alpha_1 \beta_1}\right)^{-\alpha_1} & \text{if} \quad x \leq v, \\ 1 - \left(1 + \frac{x - v^*}{\alpha_2 \beta_2}\right)^{-\alpha_2} & \text{if} \quad x > v, \end{cases} \tag{3}$$

where v is a fixed (typically unknown) constant and v^* a quantity given by

$$v^* = v - \beta_2 \alpha_2 \left(p^{-1/\alpha_2} - 1\right), \quad p = \mathsf{P}(X > v) = \left(1 + \frac{v}{\alpha_1 \beta_1}\right)^{-\alpha_1}.$$

We furthermore assume that $0 < \alpha_2 < \alpha_1$ and $\beta_i > 0$, $i = 1, 2$. As $F_X(x)$ coincides with a *generalized Pareto distribution* (GPD) with tail-index α_2 as soon as $x \geq v$, we have that F_X satisfies (2) with $\alpha_k = \alpha_2$. This means that even if v is very high and α_1 greater than 1, one enters the dangerous infinite-mean zone as soon as $\alpha_2 < 1$. VaR corresponding to F_X can be calculated explicitly,

$$\mathrm{VaR}_\gamma(X) = \begin{cases} \alpha_1 \beta_1 \left((1-\gamma)^{-1/\alpha_1} - 1\right) & \text{if} \quad \gamma \leq 1 - p, \\ v^* + \alpha_2 \beta_2 \left((1-\gamma)^{-1/\alpha_2} - 1\right) & \text{if} \quad \gamma > 1 - p. \end{cases} \tag{4}$$

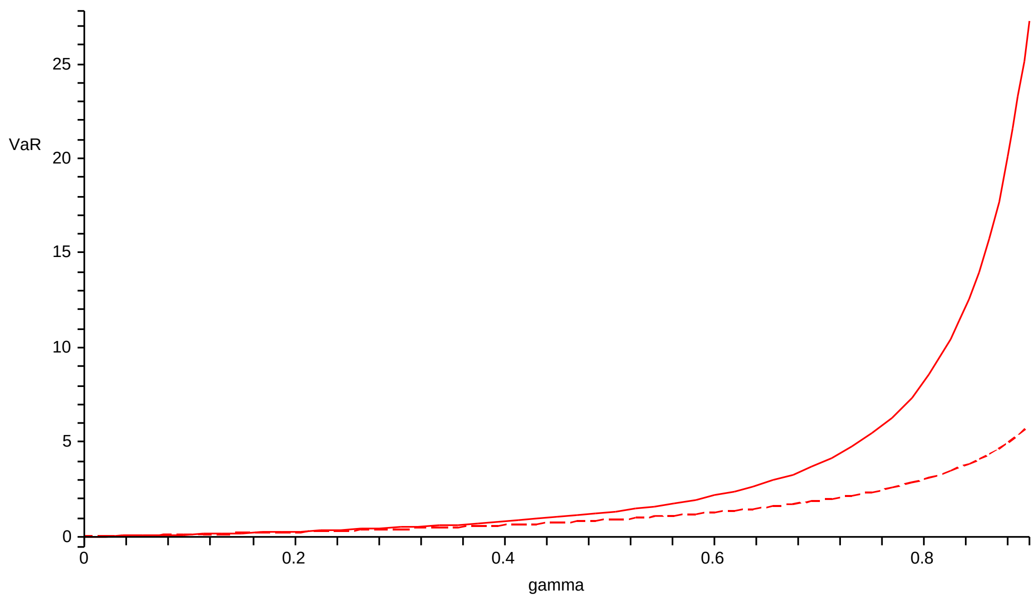

Figure 3: VaR_γ calculated via (5) with $u = 0$, $\beta_2(u) = 1$ and $\alpha_2 = 2/3$ (full line) and VaR_γ calculated via (4) with $p = 0.1$, $\beta_1 = 1$ and $\alpha_1 = 1.4$ (dashed line).

Implications of (4) are twofold. For one, any estimator of the tail-index parameter of F_X which is less than α_2 leads to an *underestimation* of VaR_γ for $\gamma > (1-p)$. This can easily happen if the threshold chosen for the (classical) EVT-based statistical inference is far below v. On the other hand however, at levels γ below $(1-p)$, VaR_γ calculated via the POT-based approximate formula

$$\mathrm{VaR}_\gamma \approx u + \beta_2(u)\alpha_2 \left((1-\gamma)^{-1/\alpha_2} - 1\right) \tag{5}$$

for some scale parameter $\beta_2(u)$ and threshold u, may lead to a vast *overestimation*. In Figure 3, we compare VaR_γ calculated via (5) and (4). The difference gets even more pronounced for large v's (or, equivalently, large values of $(1-p)$). A large v also implies that the levels γ which are of interest to the operational risk modeler possibly lie below $(1-p)$.

Mixture models of the above type can however turn out to be difficult to detect if one does not look for them. For small thresholds v, samples drawn from F_X are likely to exhibit sufficiently many points above v and the change of behavior can be detected by the mean excess plot, for instance. Large values of v however typically lead to cases where there are only very few excesses above v in the sample (10, say). Consequently, the mean excess plot may not reveal

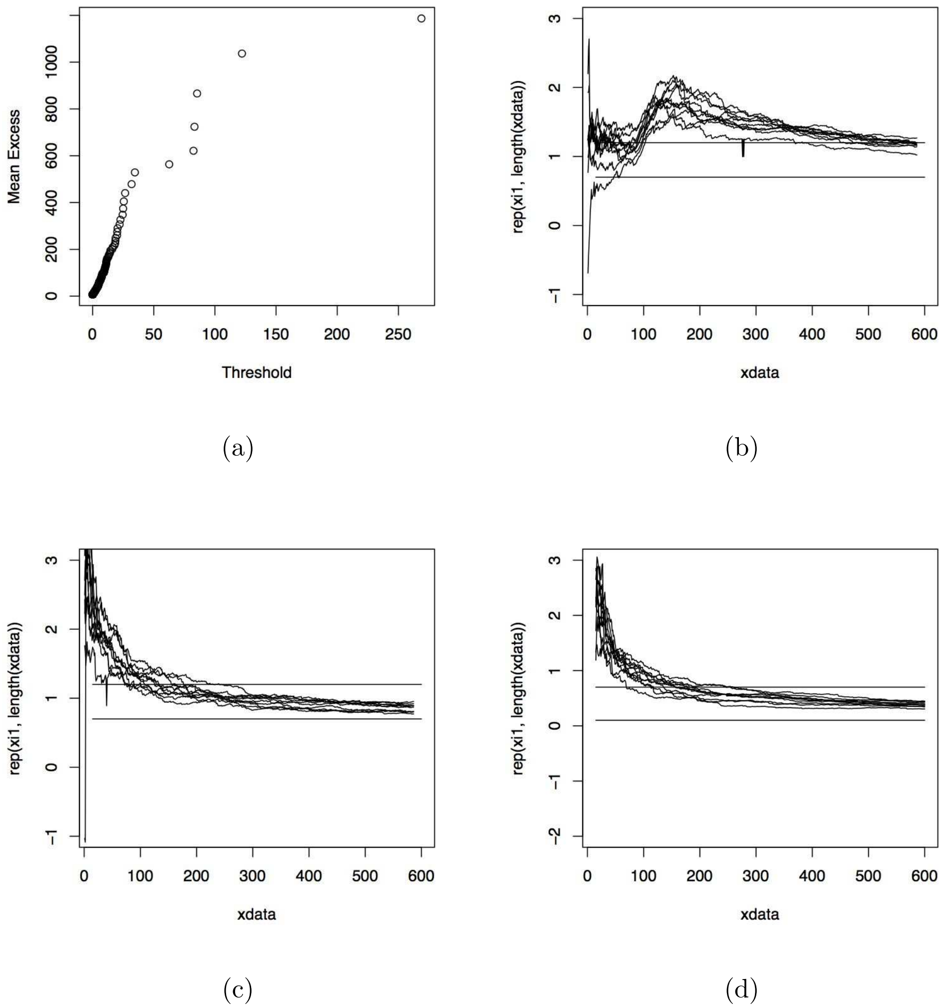

Figure 4: Classical POT model in the presence of contamination above a high threshold: mean excess plot (a) and shape plots (b-d) for 20 simulated samples with 100 contaminated data points (b) and 10 contaminated data points (c-d). The straight lines in plots (b-d) denote $1/\alpha_1$ and $1/\alpha_2$, respectively.

anything suspicious (see Figure 4 (a), further described below) and yet the points lying above v can have substantial influence on the outcomes of the statistical inference. In particular, estimates of the tail-index parameter well below 1 may result, even if α_1 is significantly larger than 1 and v is high. In Figure 4 (a-c), we simulate from the above model (3) with $\alpha_1 = 1.42$, $\alpha_2 = 0.83$. In plots (a) and (c), we take $p = 0.99$ and simulate samples with exactly 10 contaminated data points; in (b) we choose $p = 0.9$ leading to samples with 100 contaminations. Apart from the mean excess plot for one simulated sample, the figure displays the so-called shape plot (estimates of $1/\widehat{\alpha}$ over a range of thresholds) for 10 simulated samples with $p = 0.99$ (c) and $p = 0.9$ (b). For a higher number of simulations, the pictures stay very much the same except that they become less transparent and hence less suitable for our illustrative purpose.

Finally, Figure 4 shows that the model (3) can furthermore be used as an exemplar on the misuse of EVT. Even for $\alpha_2 > 1$, excesses above high thresholds have *finite mean* and yet, if we choose $\alpha_2 = 1.42$, $\alpha_1 = 10$ and $p = 0.99$ leading to ca. 10 contaminated data points, the shape plot displayed in Figure 4 (d) behaves very much like that in (c). In other words, there exist threshold choices still well in the data which produce estimates of the tail index well below 1.

Although F_X is extremely heavy tailed if $\alpha_2 < 1$, it is still “less dangerous” in terms of VaR than an exact Pareto distribution with parameter α_2, say. This example once again points out that it is crucial to gain the best possible knowledge of the data generating mechanism and not to use EVT without careful considerations. This task involves not only mathematical tools but also background information on the (especially large) data points. Mathematical tools that can help are comparisons between the classical POT model and a likelihood-based inference derived from the above model (3). Note that if $\alpha_1 = \alpha_2$, X follows a GPD distribution. This means that the classical POT model can be viewed as a special case and likelihood ratio tests can be performed for comparison. ■

In the above example, contaminated data appeared solely above some high threshold v. In practice, this would mean that in a sample of n points, only the k *largest* observations are contaminated whereas the remaining $n - k$ are not. In the next example, we generalize this idea in that we consider a model where there are k *out of* n data points contaminated.

γ	$\text{VaR}_\gamma(F_X)$	$\text{VaR}_\gamma(\text{Pareto}(\alpha_2))$	α^*
0.9	6.39	46.42	1.24
0.95	12.06	147.36	1.2
0.99	71.48	2154.43	1.08
0.999	2222.77	10^5	0.89
0.9999	10^5	$4.64 \cdot 10^6$	0.79
0.99999	$4.64 \cdot 10^6$	$2.15 \cdot 10^8$	0.75

Table 1: Value-at-Risk for mixture models.

Example 2.2. Mixture models. The model given by (3) can be viewed as a special mixture, i.e. one can show that

$$F_X = (1-p)F_1 + pF_2, \tag{6}$$

where F_2 is a (shifted) GPD distribution with parameters α_2 and β_2 and F_1 is a GPD distribution with parameters α_1 and β_1 truncated at v. In this example, we consider general mixtures, i.e. loss distributions of the form (6). Such models would typically arise when the losses are caused by one of two independent effects, each happening with a certain probability. Another example are distributions of aggregated losses, where the underlying loss process is a sum of two independent and homogeneous compound Poisson processes, a situation which we discuss more in detail in the following section.

For an illustration, consider a mixture distribution (6) with F_i exact Pareto with parameter α_i, $i = 1, 2$, i.e. $F_i(x) = 1 - x^{-\alpha_i}$ for $x \geq 1$. We furthermore assume that $0 < \alpha_2 < \alpha_1$. As in the preceding Example 2.1, it follows that

$$1 - F_X(x) = x^{-\alpha_2} h(x)$$

for h slowly varying. In other words, regardless α_1 and p, *very* high losses are driven essentially by F_2. The question how high *very high* is, is however crucial: depending on p and α_1, the asymptotics may become significant at levels which are far beyond those of interest in practice and often beyond the range of the available data. The Value-at-Risk corresponding to (6) is in general no longer given by a simple formula as in Example 2.1, but has to be calculated numerically. One can however at least gain some insight into the role of the parameters p, α_1

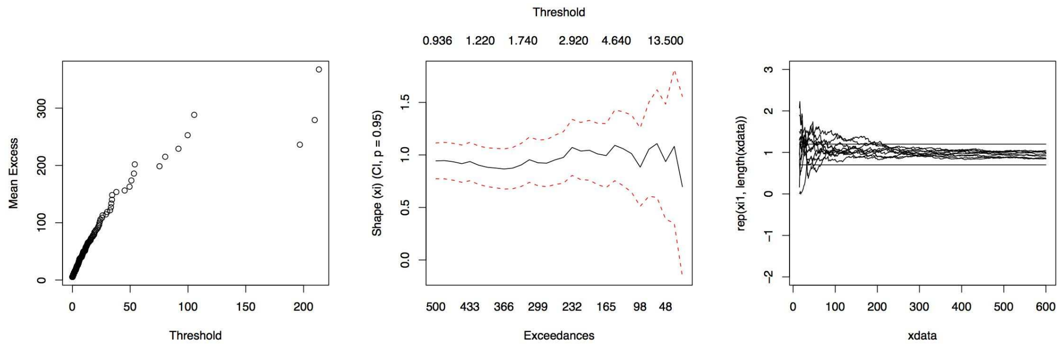

Figure 5: Classical POT model in the presence of contamination: mean excess plot (left), shape plot (middle) and shape plot for 20 simulated samples (right).

and α_2 by showing that $\mathrm{VaR}_\gamma(F_X) = \mathrm{VaR}_{\gamma_1}(F_1) = \mathrm{VaR}_{\gamma_2}(F_2)$, where γ_1 is the solution of

$$(1-\gamma_1)^{-1/\alpha_1} = \left(1 - \frac{\gamma - (1-p)\gamma_1}{p}\right)^{-1/\alpha_2}$$

and $\gamma_2 = (\gamma - (1-p)\gamma_1)/p$. Note that this in particular implies $\gamma_1 \geq \gamma$ and $\gamma_2 \leq \gamma$. Hence VaR_γ equals the Value-at-Risk corresponding to F_2, but at a level which is *lower* than γ. In Table 1, we report the values of VaR_γ corresponding to F_X given by

$$F_X(x) = 0.9(1 - x^{-1.4}) + 0.1(1 - x^{-0.6})$$

and compare them with VaR_γ of an exact Pareto distribution with tail-index $\alpha_2 = 0.6$. The difference is striking. At first glance, one may still hope that there exists an α^*, such that $\mathrm{VaR}_\gamma(F_X)$ would be consistent with VaR_γ of a Pareto distribution with tail-index α^*. Unfortunately, the fourth column of Table 1 shows that such hopes are dead-end. Also note that VaR_γ of a Pareto distribution with α_1 would lead to considerable *underestimation*, in particular at high levels γ. Worse, Table 1 also implies that, in general, the use of the classical EVT-POT model cannot give correct estimations of high quantiles. Worse still, general mixtures are far more difficult to detect than models considered in Example 2.1. Especially for small values of p and small data sets, it is fairly likely that only few points in the sample would actually come from the more dangerous distribution F_2 and one can be easily seduced by the relatively good fit of the classical POT model. Even for a comparatively high proportion of contaminated data, the fit can be reasonable however. In Figure 5 we simulate a sample of $n = 1000$ points from the above mixture model (6) with $p = 0.33$ and F_i generalized Pareto with $\alpha_1 = 1.42$, $\alpha_2 = 0.83$ and $\beta_1 = \beta_2 = 1$. Neither the mean excess plot nor the shape plot look odd. However, the plot in Figure 5 (right) reveals how misleading the classical POT model can be. Here we display shape plots of 10 simulated samples with the above parameters (as in Example 2.1, we use 10 samples merely for a better visibility). In most cases, the shape plot looks fairly straight which in the classical POT model indicates a good fit. However, the resulting estimate of α (around 1 in this case) typically lies between α_1 and α_2 which leads to overestimation of VaR at lower levels and underestimation at high levels; see the right column of Table 1. Finally, note that even if α_2 is small, the largest observations are not necessarily all contaminated. As in the preceding example, one can come up with extensions of the classical POT model which allow for this type of contamination. This would enable MLE estimation of α_1, α_2 and p as well as testing against the classical (non-contaminated) model. ■

So far, we gave examples which merely highlight potential shortcomings of the classical POT model in the presence of contamination. Further work on statistical inference for mixture models is necessary before extensions of classical EVT can be discussed in greater detail and applied to real and/or simulated operational risk data samples. Only further analyses of operational risk data will reveal whether the issue of contamination is relevant. From an understanding of some of the data, we very much believe it is. But even if the results would indeed point in this direction, the question how to determine an appropriate risk capital for very heavy-tailed risks however still stands. So far we can only say that more statistical work on these issues is badly needed. In the next section, we turn to the issue of correlation and diversification and its implications for the capital charge calculation.

3 The AMA capital charge revisited

The estimation of $\mathrm{VaR}_\gamma(L)$ needs a *joint* model for the loss random vector $(L_1, \ldots, L_d)$, i.e. concrete assumptions on the interdependence between the L_k's have to be made complementary to (2), say. Because of a lack of specific distributional models for the vector $(L_1, \ldots, L_d)$, the Basel II AMA guidelines leading to the Loss Distribution Approach (LDA) suggest to use $\sum_{k=1}^{d} \mathrm{VaR}_{99.9\%}(L_k)$ for a capital charge and allow for a diversification reduction of the latter under appropriate correlation assumptions. These assumptions have to be made explicit as well as plausible to the local regulator.

This is exactly the point where we want to voice a *warning*. If we accept the results of the extensive and careful statistical analysis in Moscadelli (2004), then the dfs of $L_1, \ldots, L_d$ satisfy (2) with α_k's less than 2 and several less than 1, hence not only are correlation coefficients not defined, but the loss dfs are extremely heavy-tailed. This may have serious consequences as we will see below. Basing a capital charge reduction on the notion of correlation may be meaningless, to say the least. The situation however gets worse in the face of extreme heavy-tailedness. To start with, one should ask whether

$$\mathrm{VaR}_\gamma(L) = \mathrm{VaR}_\gamma\left(\sum_{k=1}^{d} L_k\right) \le \sum_{k=1}^{d} \mathrm{VaR}_\gamma(L_k); \tag{7}$$

this is referred to as *subadditivity* of Value-at-Risk and is one of the crucial properties of the notion of *coherence* as discussed in Artzner et al. (1999). First note that equality in (7), i.e.

$$\mathrm{VaR}_\gamma\left(\sum_{k=1}^{d} L_k\right) = \sum_{k=1}^{d} \mathrm{VaR}_\gamma(L_k) \tag{8}$$

is obtained under the assumption of *comonotonicity* of the vector $(L_1, \ldots, L_d)$. This means that there exists a rv Z and increasing deterministic functions $f_1, \ldots, f_d$ so that for $k = 1, \ldots, d$, $L_k = f_k(Z)$. For ease of notation, take $d = 2$. Then whenever L_1 and L_2 have finite second moments, (8) holds if the correlation coefficient $\varrho(L_1, L_2)$ is *maximal*; this situation is referred to as *perfect dependence* (or *perfect correlation*). The value of this maximal correlation $\varrho_{\max}$ is possibly – even typically – less than 1. As an example, Embrechts et al. (2002) calculate $\varrho_{\max}$ in the case of two lognormal risks. They also show in that example that $\varrho_{\max}$ can be made as close to zero as one likes *keeping* comonotonicity. In the following example, we examine $\varrho_{\max}$ and $\varrho_{\min}$ for Pareto risks as in (2).

Example 3.1. Bounds on correlation for Pareto risks. Consider L_1 and L_2 exact Pareto with parameter α and β respectively, i.e. take $\mathsf{P}(L_1 > x) = x^{-\alpha}$ and $\mathsf{P}(L_2 > x) = x^{-\beta}$ for $x \ge 1$. Recall that the correlation between L_1 and L_2 is defined if and only if both α and β are strictly greater than 2. Given that assumption, $\varrho(L_1, L_2)$ is maximal if L_1 and L_2 are comonotonic. Similarly, $\varrho(L_1, L_2)$ attains its smallest value for L_1 and L_2 *countermonotonic*, i.e. if there exists an increasing function f_1, a decreasing function f_2 and a rv Z so that $L_i = f_i(Z)$, $i = 1, 2$. The maximum and minimum possible correlation can be calculated

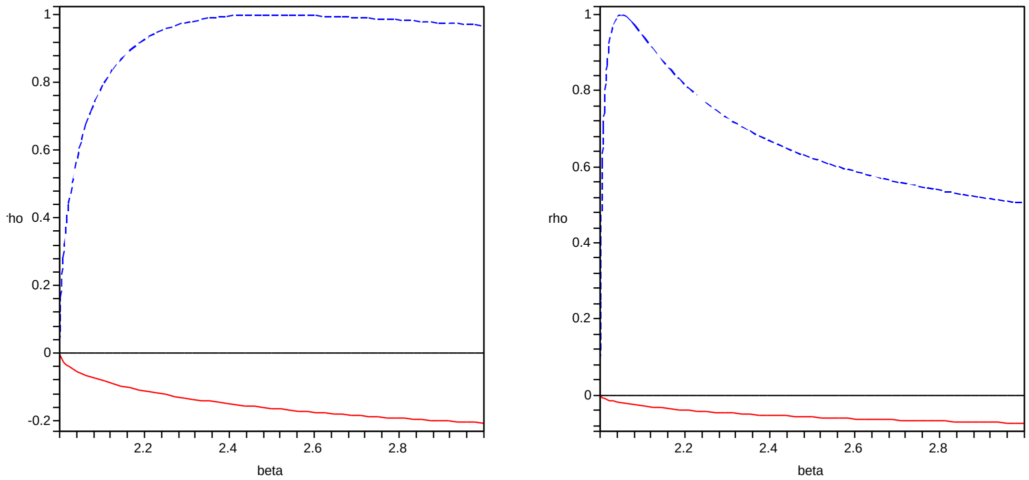

Figure 6: Maximum (dashed line) and minimum (full line) attainable correlation as a function of β for $L_1 \sim$ Pareto(2.5) (left picture), $L_1 \sim$ Pareto(2.05) (right picture) and $L_2 \sim$ Pareto(β).

explicitly and one obtains that

$$\varrho_{\max}(L_1, L_2) = \frac{\sqrt{\alpha\beta(\alpha-2)(\beta-2)}}{\alpha\beta - \alpha - \beta}, \tag{9}$$

$$\varrho_{\min}(L_1, L_2) = \frac{\sqrt{(\alpha-2)(\beta-2)}\left((\alpha-1)(\beta-1)B\left(1-\frac{1}{\alpha}, 1-\frac{1}{\beta}\right) - \alpha\beta\right)}{\sqrt{\alpha\beta}}, \tag{10}$$

where $B(x, y)$ is the Beta function given by $B(x, y) = \int_0^1 u^{x-1}(1-u)^{y-1}du$. It is clear from (9) and (10) that the bounds on ϱ are not equal to 1 and -1 except asymptotically or in special cases. First, note that $\varrho_{\max} = 1$ if and only if $\alpha = \beta$. Moreover, $\varrho_{\min} > -1$ as can easily be checked numerically; it follows immediately from a more general result stated in Proposition 5.2.7 of Denuit et al. (2005). For β fixed, we further have that $\lim_{\alpha\to\infty} \varrho_{\max} = \sqrt{\beta(\beta-2)}/(\beta-1)$ and $\lim_{\alpha\to\infty} \varrho_{\min} = \sqrt{\beta(\beta-2)}(-\psi_0(1)+\psi_0(2-1/\beta)-1)$ where ψ_0 denotes the digamma function, which further implies $\lim_{(\alpha,\beta)\to(\infty,\infty)} \varrho_{\max} = 1$ and $\lim_{(\alpha,\beta)\to(\infty,\infty)} \varrho_{\min} = 1-\pi^2/6 \doteq -0.645$. On the other hand, both $\varrho_{\max}$ and $\varrho_{\min}$ tend to zero as soon as one of the parameters converges to 2 (i.e. if either L_1 or L_2 becomes increasingly heavy-tailed), as illustrated in Figure 6. This means that heavy-tailed Pareto risks with tail indexes close to 2 will typically exhibit only very small correlation. ■

We next give a historical example, which shows that (8) is not restricted to comonotonicity only. In some of the examples below, we use two-sided distributions in order to keep the historical character; this will however *not* affect our practical conclusions.

Example 3.2. The Cauchy distribution. Consider L_1 and L_2 independent standard Cauchy, i.e. the density functions are

$$f_{L_k}(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R}, \quad k = 1, 2.$$

For L_1, L_2 independent, we have that $(L_1 + L_2)/2$ is again standard Cauchy. Hence,

$$\mathsf{P}(L_1 + L_2 > 2x) = \mathsf{P}(X_1 > x)$$

so that, for any given γ,

$$\mathrm{VaR}_\gamma(L_1 + L_2) = 2\,\mathrm{VaR}_\gamma(L_1) = \mathrm{VaR}_\gamma(L_1) + \mathrm{VaR}_\gamma(L_2)$$

as in the comonotonic case (8), but this time under the assumption of *independence*. As communicated to us by Professor Benoit Mandelbrot, Cauchy realized the "diversification fallacy" underlying this property. Note that for the Cauchy case, $\mathrm{E}(|L_k|) = \infty$. There are several, for our paper relevant, quotes on Augustin-Louis Cauchy to be found in Mandelbrot and Hudson (2004). For instance, on p. 39:

> "The difference between the extremes of Gauss and Cauchy could not be greater. They amount to two different ways of seeing the world: one in which big changes are the result of many small ones, or another in which major events loom disproportionately large. 'Mild' and 'wild' chance are my generalizations from Gauss and Cauchy."

On p. 37, the same authors introduce the Cauchy distribution as a model of the blindfolded archer's score in which case "the largest shot will be nearly as large as the sum of all the others. One miss by a mile completely swamps 100 shots within a few yards of the target." We come back to this issue in Section 4, using the notion of subexponentiality. ■

In the next example we look at the extremely heavy-tailed case more in detail.

Example 3.3. Infinite mean Pareto models. This example is taken from Embrechts et al. (2002) (see Examples 6 and 7) and has important consequences concerning the notion of diversification in economics as discussed in Ibragimov (2005). Take L_1 and L_2 independent and identically distributed with $\mathsf{P}(L_1 > x) = \mathsf{P}(L_2 > x) = x^{-1/2}$, $x \geq 1$, i.e. L_1 and L_2 are extremely heavy-tailed with infinite mean. In this case, one shows that

$$\mathsf{P}(L_1 + L_2 \leq x) = 1 - \frac{2\sqrt{x-1}}{x} < \mathsf{P}(2L_1 \leq x)$$

for $x > 2$. It follows that

$$\mathrm{VaR}_\gamma(L_1 + L_2) > \mathrm{VaR}_\gamma(2L_1) = \mathrm{VaR}_\gamma(L_1) + \mathrm{VaR}_\gamma(L_2)$$

so that from the point of view of VaR_γ, independence is *worse* than perfect dependence (to be interpreted as comonotonicity). There is nothing special about the choice 1/2 as power exponent in the df of L_1 and L_2, the same result holds whenever

$$\mathsf{P}(L_k > x) = x^{-\alpha}, \quad x \geq 1, \tag{11}$$

for $0 < \alpha < 1$. Klüppelberg and Rootzén (1999) arrive at the same conclusion that "big is not always beautiful" using an asymptotic approximation for $\mathrm{VaR}_\gamma(L_1+L_2)$. The case $\alpha = 1$ (like the Cauchy case) is also discussed in Denuit and Charpentier (2004), p. 236. In the words of Ibragimov (2005), "Many economic models are robust to heavy-tailedness assumptions as long as the distributions entering these assumptions are not extremely heavy-tailed (i.e. $\alpha > 1$). But the implications of these models are reversed for distributions with extremely heavy tails (i.e. $0 < \alpha \leq 1$)." As reported in Example 7 of Embrechts et al. (2002), whenever $\alpha > 1$ in (11), and even more generally in (2), and γ is large enough, VaR_γ is subadditive (i.e diversification holds) for iid risks. ■

The reader may ask why the transition from finite to infinite mean causes the problem. One fundamental reason is the Strong Law of Large Numbers (SLLN) for which the existence of the first moment is a necessary (as well as sufficient) condition. To shed more light on this issue, we first restate the following result from Embrechts et al. (1997), Theorem 2.1.5.

Theorem 1. Marcinkiewicz-Zygmund SLLN. *Suppose that $X_1, X_2, \dots$ are iid random variables and denote $S_n = \sum_{k=1}^n X_k$. For $p \in (0, 2)$, the SLLN*

$$n^{-1/p}(S_n - an) \to 0 \quad a.s. \tag{12}$$

holds for some real constant a if and only if $\mathrm{E}(|X|^p) < \infty$. If (X_n) obeys the SLLN (12), then we can choose

$$a = \begin{cases} 0 & \text{if} \quad p < 1, \\ \mu = \mathrm{E}(X_1) & \text{if} \quad p \in [1, 2). \end{cases}$$

Moreover, if $\mathrm{E}(|X|^p) = \infty$ *for some* $p \in (0,2)$*, then for every real* a*,*

$$\limsup_{n\to\infty} n^{-1/p}|S_n - an| = \infty \quad a.s. \tag{13}$$

From (12) it follows that for a portfolio of iid risks with finite mean μ, $\mathrm{VaR}_\gamma(S_n)$ (not compensated for the expected losses) grows like μn. Since for all reasonable risk distributions and sufficiently high confidence levels γ, $\mathrm{VaR}_\gamma(X_1) > \mu$, we obtain (asymptotic) subadditivity. Diversification works, and this is also the main reason that the classical SLLN (12) (with $\mu < \infty$) forms the methodological backbone of premium calculations in insurance. The situation however becomes very different whenever $\mu = \infty$, i.e. we fall into the (13) regime. Take for instance $X_1 \geq 0$ so that $\mathsf{P}(X_1 > x) = x^{-\alpha}h(x)$ for some slowly varying function h and $0 < \alpha < 1$. Then for some $\varepsilon > 0$, $p = \alpha + \varepsilon < 1$ and $\mathrm{E}(X_1^p) = \infty$ so that (13) holds for that value of p. As a consequence, $\mathrm{VaR}_\gamma(S_n)$ grows at least like $n^{1/p} > n$ leading to the violation of subadditivity (asymptotically). In other words, diversification no longer holds. There is much more to be said on the interplay between the (S)LLN and diversification; see for instance Samuelson (1963), Malinvaud (1972) and Haubrich (1998). ■

In the next example, we combine the above findings generalizing the Cauchy case discussed in Example 3.2. Once more, the fundamental split between $\mu < \infty$ (i.e. diversification holds) and $\mu = \infty$ (i.e. the non-diversification situation) will follow clearly.

Example 3.4. α-Stability. Recall the class of α-stable distributions; see Embrechts et al. (1997), Section 2.2 and in particular p. 78. If $L_1, \ldots, L_d$ are independent symmetric α-stable, then

$$d^{-1/\alpha}(L_1 + \cdots + L_d) \stackrel{\mathcal{D}}{=} L_1, \tag{14}$$

where $\stackrel{\mathcal{D}}{=}$ means equal in distribution. The Cauchy case in Example 3.2 corresponds to $\alpha = 1$ and $d = 2$. An immediate consequence of (14) is that for all γ,

$$\mathrm{VaR}_\gamma\left(\sum_{k=1}^{d} L_k\right) = d^{1/\alpha}\,\mathrm{VaR}_\gamma(L_1) = d^{1/\alpha-1}\left(\sum_{k=1}^{d}\mathrm{VaR}_\gamma(L_k)\right)$$

with

$$d^{1/\alpha-1}\begin{cases} >1 & \text{if} \quad \alpha < 1, \\ =1 & \text{if} \quad \alpha = 1, \\ <1 & \text{if} \quad \alpha > 1. \end{cases}$$

In this example, it is clearly seen how the change from a finite mean model ($\alpha > 1$) with subadditivity for VaR moves to superadditivity in the infinite mean case ($\alpha < 1$). The example can be further detailed restricting the L_k's to be positive and replacing "symmetric α-stability" by "being in the domain of attraction of an α-stable law" which leads to models like (2); see Embrechts et al. (1997), Definition 2.2.7 and Theorem 2.2.8. A similar example is also considered in Klüppelberg and Rootzén (1999); they give a nice discussion on the inability of VaR to risk manage catastrophic events. For an early discussion of this example in the economic literature, see also Fama and Miller (1972), p. 271. ■

The basic implication of the previous examples is that for extremely heavy-tailed loss dfs, i.e. (2) with $0 < \alpha_k < 1$, and independent L_k's,

$$\mathrm{VaR}_\gamma(L_1 + \cdots + L_d) > \sum_{k=1}^{d} \mathrm{VaR}_\gamma(L_k),$$

at least for γ sufficiently large. This raises the issue how to justify capital reduction due to "diversification". The quotes are included to hint at the need for a definition of the concept of diversification in this case; see for instance Tasche (2005) for a relevant discussion and Embrechts et al. (2002), Remark 2. To what extent this is merely a theoretical issue needs settling; operational risk managers however have to be aware of the fact that in the presence of extremely heavy-tailed loss distributions, standard economic thinking may have to be reconsidered. Only more extensive data analyses will reveal the extent to which the above is relevant. For now we have to accept that a VaR-based risk capital calculation for operational risk is not yet fully understood.

A natural question one might ask in the light of the problems encountered with VaR in the presence of extremely heavy-tailed distributions ($\alpha < 1$), is whether or not a different risk measure exists that would be coherent (i.e. subadditive) and hence for which the diversification argument would still hold. The natural candidate, the *expected shortfall*

$$\mathrm{ES}_{99.9\%} = \mathrm{E}\left(L_k | L_k > \mathrm{VaR}_{99.9\%}\right),$$

is not defined for $\alpha_k < 1$; indeed in this case $\mathrm{E}(L_k) = \infty$. Moreover, it follows from Delbaen (2002), Theorem 13, that *any* coherent risk measure ρ which only depends on the distribution function of the risks and for which $\rho \geq \mathrm{VaR}_\gamma$ holds *has* to satisfy $\rho \geq \mathrm{ES}_\gamma$. The expected shortfall is therefore the *smallest* such coherent risk measure larger than VaR. Formulated differently: there exists *no* distribution invariant coherent risk measure larger than VaR which yields finite risk capital for infinite mean Pareto distributions. This leaves little hope for finding an easy solution for the capital charge problem for portfolios with extremely heavy tails. This hope is further shattered by a result from functional analysis; see Rudin (1973), Section 1.47. In the spaces $\mathbb{L}^p = \{X \text{ rv} : \mathrm{E}\,|X|^p < \infty\}$ with $0 < p < 1$ (corresponding to our $0 < \alpha < 1$ case) there exist *no* convex open sets other than the empty set and $\mathbb{L}^p$ itself. As a consequence, 0 is the only continuous linear functional on $\mathbb{L}^p$[1]. Translated into our terminology: any systematic and reasonable handling of infinite mean risks will be very challenging indeed.

We finally want to bring the interesting papers Wüthrich (2003) and Alink et al. (2004) to the operational risk modeler's attention. In these papers, the authors show that "for d identically and continuously distributed *dependent* risks $L_1, \ldots, L_d$, the probability of a large aggregate loss of $\sum_{k=1}^{d} L_k$ scales like the probability of a large individual loss of L_1, times a *proportionality factor* q_d." They show that the value of q_d, and hence the resulting (non)-subadditivity of VaR, depends on the interplay between the interdependence properties of $L_1, \ldots, L_d$ and the heavy-tailedness of the L_k's. Further papers containing results for dependent risks are Ibragimov (2005), Section 17, and Daníelsson et al. (2005).

4 The one loss causes ruin problem

In this section we present some further tools and results, mainly from the world of insurance mathematics, which in the light of the previous discussions may be useful in understanding the observed properties of operational risk data.

We first make some comments about loss portfolios where the (iid, say) losses follow a Pareto-type distribution as in (2) with tail-index α. Based on the concept of Lorenz curve in economics as discussed in Embrechts et al. (1997), Section 8.2, a large claim index is introduced explaining what percentage of the individual losses constitutes a certain percentage of the total portfolio loss. For instance, the famous $20 - 80$ rule corresponds to $\alpha = 1.4$. This means that, in an iid Pareto portfolio with tail-index 1.4, 20% of the individual losses produce 80% of the total portfolio loss. In the light-tailed exponential case, we roughly have a $50 - 80$ rule. In contrast, $\alpha = 1.01$ (a model with still finite mean, but only just) leads to a $0.1 - 95$ rule, i.e. 0.1% of

[1] Rudin states, "This is, of course, in *violent* contrast to the familiar case $p \geq 1$."

the losses is responsible for 95% of the total loss amount; compare this with the comments by Mandelbrot and Hudson quoted in Example 3.2. In such models (and definitely for $\alpha < 1$) we enter the "one loss causes ruin" regime as discussed in Asmussen (2000), p. 264, as the "one large claim" heuristics. See also Figure 1.3.7 in Embrechts et al. (1997) for a simulated illustration of this phenomenon in a ruin model context. A discussion of this figure and its consequences is also to be found in Mandelbrot and Hudson (2004), p. 232.

There are various ways in which the "one loss causes ruin" paradigm manifests itself. In the context of operational risk, the route via subexponentiality is a very natural one; see Embrechts et al. (1997), Section A 3.2. Take $X_1, \ldots, X_n$ positive iid random variables with common distribution function F_X, denote $S_n = \sum_{k=1}^n X_k$ and $M_n = \max(X_1, \ldots, X_n)$. The distribution function F_X is called *subexponential* (denoted by $F_X \in \mathcal{S}$) for some (and then for all) $n \geq 2$ if

$$\lim_{x \to \infty} \frac{\mathsf{P}(S_n > x)}{\mathsf{P}(M_n > x)} = 1, \tag{15}$$

i.e. a large value of S_n is mainly determined by a single large individual value M_n; see again the final comment in Example 3.2. Examples of distributions satisfying (15) are Pareto-type distributions like in (2), the lognormal and the loggamma distributions for instance. Restated in the case of L in (15) this yields that for iid business line and/or risk type rvs $L_1, \ldots, L_d$ with $F_{L_i} \in \mathcal{S}$ one has that for x large,

$$\mathsf{P}(L > x) \sim \mathsf{P}(\max(L_1, \ldots, L_d) > x).$$

Equivalently one can write $\mathsf{P}(L > x) \sim d\mathsf{P}(L_1 > x)$ for x large.

A further interesting property of subexponential distributions, relevant for operational risk, reveals itself when we reconsider L_k in its frequency-severity decomposition, i.e. if we assume that for each $k = 1, \ldots, d$,

$$L_k = \sum_{i=1}^{N_k} X_i(k) \tag{16}$$

for iid loss severity rvs $X_i(k)$, $i \geq 1$, independent of the loss frequency rv N_k, $k = 1, \ldots, d$. In this case, the rvs L_k in (16) are referred to as *compound* rvs. We assume that the $X_i(k)$'s have distribution function $F_{X(k)}$, $k = 1, \ldots, d$. Several useful results now hold; see Embrechts et al. (1997) for details. First of all, if for some $\varepsilon > 0$, $\sum_{n=1}^{\infty} (1+\varepsilon)^n \mathsf{P}(N_k = n) < \infty$ (satisfied for instance in the important binomial, Poisson and negative binomial cases) and $F_{X(k)} \in \mathcal{S}$, then

$$\lim_{x\to\infty} \frac{\mathsf{P}(L_k > x)}{1 - F_{X(k)}(x)} = \mathrm{E}(N_k). \tag{17}$$

This implies that for large x, the tail of the aggregate loss distribution, $\mathsf{P}(L_k > x)$, is mainly determined by the tail of one individual loss. Hence $\mathsf{P}(L_k > x) \sim \mathrm{E}(N_k)\mathsf{P}(X_i(k) > x)$ for x large. In particular, if $\mathsf{P}(X_i(k) > x) = x^{-\alpha_k} h_k(x)$ with h_k slowly varying, then $\mathsf{P}(L_k > x) \sim \mathrm{E}(N_k)x^{-\alpha_k} h_k(x)$ for $x \to \infty$. It is interesting to note that the converse also holds, i.e. if $\mathsf{P}(L_k > x)$ has power-tail behavior, then so does the individual loss df $\mathsf{P}(X_i(k) > x)$. The property $\mathsf{P}(L_k > x) \sim \mathrm{E}(N_k)\mathsf{P}(X_i(k) > x)$ for $x \to \infty$ holds if and only if $F_{X(k)} \in \mathcal{S}$; a proof of this important result in the compound Poisson case is to be found in Embrechts et al. (1997), Theorem A.3.19. We want to stress that asymptotic formulas like (17) are of methodological rather than numerical importance. For instance, a methodological consequence of (17) is that the so-called maximum domain of attraction conditions for the use of EVT are equivalently satisfied (or not) by the individual loss dfs $F_{X(k)}$ and the aggregated loss dfs $\mathsf{P}(L_k \leq x)$. The fact that estimates based on (17) may lead to unsatisfactory numerical results is well documented; see for instance Rolski et al. (1998), p. 176, and De Vylder and Goovaerts (1984).

For the total sum $L = \sum_{k=1}^{d} L_k$, results similar to (17) hold under extra conditions. One such example is the *compound Poisson* case. Suppose that the frequency rvs N_k in (16) are independent and Poisson(λ_k) distributed for $k = 1, \ldots, d$ and further that the severity rvs $X_i(k)$ are independent for $k = 1, \ldots, d$. Hence $L_1, \ldots, L_d$ are independent and one easily shows that $L = \sum_{k=1}^{d} L_k$ is compound Poisson as well, with intensity $\lambda = \sum_{k=1}^{d} \lambda_k$ and loss severity df $F = \sum_{k=1}^{d} \frac{\lambda_k}{\lambda} F_{X(k)}$. Note that F is a mixture distribution as discussed in Example 2.2. The "one loss causes ruin" paradigm in this case translates into the fact that the individual loss df $F_{X(k)}$, $k = 1, \ldots, d$, with the heaviest tail determines the tail of the distribution of the total loss L. Suppose that the *individual* loss dfs $F_{X(k)}$, $k = 1, \ldots, d$, satisfy

$$\mathsf{P}(X(k) > x) = x^{-\alpha_k} h_k(x),$$

with $\alpha_1 < \alpha_2 < \cdots < \alpha_d$, then in the above compound Poisson case,

$$\lim_{x\to\infty} \frac{\mathsf{P}(L > x)}{\mathsf{P}(X(1) > x)} = \lambda_1, \tag{18}$$

so that a very high loss for the total sum $L = \sum_{k=1}^{d} L_k$ is essentially due to a very high value for an *individual* loss in the most heavy-tailed business line (say). Practitioners are well aware of this phenomenon: it is the very few largest losses that cause the main concern. This is of particular importance in the context of operational risk. The above results hopefully give some methodological support for this practical awareness. The results stated can also be formulated for subexponential dfs and more general counting random variables N_k, $k = 1, \ldots, d$. For details, see Embrechts et al. (1997), McNeil et al. (2005) and the references therein.

5 Conclusion

In this paper, we have presented some issues underlying the AMA-LDA modeling of operational risk. The key message addressed the need for a very careful handling of quantitative risk measurement of extremely heavy-tailed loss data. The transition from finite to infinite mean models marks a risk management step into a methodological danger zone. The extent to which operational risk data really exhibit such heavy-tailed behavior needs further analysis. It is therefore to be hoped that industry will share (some of) its data with academia. In the mean time, we can just warn aginst a possibly too naive optimism concerning the calculation of an operational risk, AMA based capital charge.

The issue is perhaps less one of finite versus infinite mean, but more one of "Do we really understand the loss-generating mechanism underlying operational risk data flowing into bank internal data warehouses or into quantitative impact studies in the hands of the regulators worldwide?" That the very large losses really drive the capital charge calculation was obvious from the start. In our paper, we aimed at clarifying some of the issues underlying this "one huge loss causes havoc" paradigm. When handled correctly, EVT based analysis offers a perfect tool for pointing at data which are about to enter this paradigm; when this happens, judgment has to be added to statistical analysis. In Section 2, we presented some models which allow to help such judgment calls: we strongly believe that data contamination is a core issue in need for further research. Even a small amount of contamination can completely destroy the good statistical properties of estimators used. Robust statistics is a framework which allows to analyze these issues in detail; see for instance Dell'Aquila and Ronchetti (2006) and Reiss and Thomas (2001). A key word to look for is *robust* EVT; Dell'Aquila and Embrechts (2006) contains a brief discussion on some of these issues.

Other fields of interesting research concern the definition of risk measures and/or risk measurement procedures to be used in the presence of very heavy-tailed data. One way toward the calculation of high quantiles (VaR_γ with γ close to 1) could be the use of finite upper limit EVT, where the upper limit is for instance set at an accounting value marking insolvency of the company. It would certainly be interesting to compare EVT estimates for the risk capital both in a finite as well as an infinite support model. A further issue which needs investigation is whether a VaR estimate at a lower confidence level, 90% say, combined with a scaling to 99.9%, could be useful in the context of operational risk. This method has been used successfully in the analysis of market risk; see for instance Kaufmann (2004) for a comprehensive discussion and some VaR-based examples in Embrechts and Hoeing (2006).

Whatever road of investigation one follows, the presence of extremely heavy-tailed data will no doubt make the work both interesting as well as challenging.

Acknowledgements

This work was partly supported by the NCCR FINRISK Swiss research program and RiskLab, ETH Zurich. The authors would like to thank the participants of the meeting “Implementing an AMA to Operational Risk” held at the Federal Reserve Bank of Boston, May 18 - 20, 2005, for several useful comments. In particular, Jason Perry is thanked for comments on the earlier version of the paper. They also thank Freddy Delbaen and Peter Zweifel for helpful discussions related to Section 2, and Rosario Dell’Aquila for discussions on robust statistics.

References

Alink, S., Löwe, M., and Wüthrich, M. V. (2004). Diversification of aggregate dependent risks. *Insurance: Mathematics and Economics*, 35:77–95.

Artzner, P., Delbaen, F., Eber, J., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9:203–228.

Asmussen, S. (2000). *Ruin Probabilities*. World Scientific, Singapore.

Chavez-Demoulin, V., Embrechts, P., and Nešlehová, J. (2006). Quantitative models for operational risk: Extremes, dependence and aggregation. *The Journal of Banking and Finance*, to appear.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer, London.

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

de Fontnouvelle, P. (2005). Results of the Operational Risk Loss Data Collection Exercise (LDCE) and Quantitative Impact Study (QIS). Presentation. *Implementing an AMA to Operational Risk*, Federal Reserve Bank of Boston, May 18-20, http://www.bos.frb.org/bankinfo/conevent/oprisk2005/.

de Fontnouvelle, P., Rosengren, E., and Jordan, J. (2004). Implications of alternative operational risk modeling techniques. Preprint, Federal Reserve Bank of Boston.

De Vylder, F. and Goovaerts, M. (1984). Bounds for classical ruin probabilities. *Insurance: Mathematics and Economics*, 3:121–131.

Delbaen, F. (2002). Coherent risk measures. Cattedra Galileiana Pisa, Scuola Normale Superiore, Publicazioni della Classe di Scienze.

Dell'Aquila, R. and Embrechts, P. (2006). Extremes and robustness: a contradiction? *Financial Markets and Portfolio Management*, to appear.

Dell'Aquila, R. and Ronchetti, E. (2006). *Robust Statistics and Econometrics with Economic and Financial Applications*. Willey, Chichester, to appear.

Denuit, M. and Charpentier, A. (2004). *Mathématiques de l'Assurance Non-Vie. Tome 1: Principes Fondamentaux de Théorie du Risque*. Economica, Paris.

Denuit, M., Dhaene, J., Goovaerts, M., and Kaas, R. (2005). *Actuarial Theory for Dependent Risks*. Wiley, Chichester.

Embrechts, P. and Hoeing, A. (2006). Extreme var scenarios in high dimensions. Preprint, ETH Zurich.

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

Embrechts, P., McNeil, A., and Straumann, D. (2002). Correlation and dependence in risk management: Properties and pitfalls. In Dempster, M., editor, *Risk Management: Value at Risk and Beyond.*, pages 176–223. Cambridge University Press, Cambridge.

Fama, E. F. and Miller, M. H. (1972). *The Theory of Finance.* Dryden Press, Hinsdale, IL.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions.* Wiley, New York.

Haubrich, J. G. (1998). Bank diversification: laws and fallacies of large numbers. *Economic Review, Issue QII*, pages 2–9.

Huber, P. J. (1981). *Robust Statistics.* Wiley, New York.

Ibragimov, R. (2005). On the robustness of economic models to heavy-tailedness assumptions. Preprint, Harvard Institute of Economic Research, Harvard University.

Kaufmann, R. (2004). *Long-Term Risk Management.* PhD thesis, ETH Zurich.

Klüppelberg, C. and Rootzén, H. (1999). A single number can't hedge against economic catastrophes. *Ambio*, 28(6):550–555.

Malinvaud, E. (1972). The allocation of individual risks in large markets. *Journal of Economic Theory*, 4:312–328.

Mandelbrot, B. (1997). *Fractals and Scaling in Finance.* Springer, New York.

Mandelbrot, B. and Hudson, R. (2004). *The (Mis)Behavior of Markets.* Basic Books, New York.

McLachlan, G. and Basford, K. (1988). *Mixture Models: Inference and Applications to Clustering.* Marcel Dekker, New York and Basel.

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

Moscadelli, M. (2004). The modelling of operational risk: experience with the analysis of the data collected by the Basel committee. Technical Report 517, Banca d'Italia.

Rachev, S. and Mittnik, S. (2000). *Stable Paretian Models in Finance.* Wiley, Chichester.

Reiss, R.-D. and Thomas, J. (2001). *Statistical Analysis of Extreme Values, 2nd edition.* Birkhäuser, Basel.

Rolski, T., Schmidli, H., Schmidt, V., and Teugels, J. (1998). *Stochastic Processes for Insurance and Finance.* Wiley, New York.

Rudin, W. (1973). *Functional Analysis.* McGraw-Hill, New York.

Samuelson, P. A. (1963). Risk and uncertainty: a fallacy of large numbers. *Scientia*, 57:1–6.

Tasche, D. (2005). Risk contibutions in an asymptotic multi-factor framework. Discussion paper, Deutsche Bundesbank. Presented at the workshop on “Concentration Risk in Credit Portfolios”, Eltville, November 18-19, 2005.

Wüthrich, M. V. (2003). Asymptotic value-at-risk estimates for sums of dependent random variables. *ASTIN Bulletin*, 33(1):75–92.