Quantitative Models for Operational Risk: Extremes, Dependence and Aggregation

V. Chavez-Demoulin, P. Embrechts and J. Nešlehová*

Department of Mathematics

ETH-Zürich

CH-8092 Zürich Switzerland

`http://www.math.ethz.ch/~valerie/`

`http://www.math.ethz.ch/~embrechts/`

`http://www.math.ethz.ch/~johanna/`

Abstract

Due to the new regulatory guidelines known as Basel II for banking and Solvency 2 for insurance, the financial industry is looking for qualitative approaches to and quantitative models for operational risk. Whereas a full quantitative approach may never be achieved, in this paper we present some techniques from probability and statistics which no doubt will prove useful in any quantitative modelling environment. The techniques discussed are advanced peaks over threshold modelling, the construction of dependent loss processes and the establishment of bounds for risk measures under partial information, and can be applied to other areas of quantitative risk management[1].

JEL classification: C.14; G.10; G.21

Keywords: Copula; Dependence; Fréchet class problems; Generalized Pareto distribution;

*Corresponding author: Johanna Nešlehová. Address: RiskLab, Department of Mathematics, ETH Zurich, Rämistrasse 101, CH-8092 Zurich, Switzerland. Tel: +41 44 632 32 27, Fax: +41-1-632 15 23, Email: johanna@math.ethz.ch

[1]This paper was presented as an invited contribution at the meeting "Implementing an AMA for Operational Risk", Federal Reserve Bank of Boston, May 18-20, 2005

Mass transportation; Operational risk; Peaks over threshold; Point process; Risk aggregation; Statistics of extremes.

1 Introduction

Managing risk lies at the heart of the financial services industry. Regulatory frameworks, such as Basel II for banking and Solvency 2 for insurance, mandate a focus on operational risk. In the Basel framework, operational risk is defined as the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events. This definition includes legal risk, but excludes strategic and reputational risk. A fast growing literature exists on the various aspects of operational risk modelling; see for instance Cruz (2002), Cruz (2004) and King (2001) for some textbook treatments. For a discussion very much in line with our paper, see Chapter 10 in McNeil et al. (2005).

In this paper we discuss some of the more recent stochastic methodology which may be useful towards the quantitative analysis of *certain types* of operational loss data. We stress the "certain types" in the previous sentence. Indeed, not all operational risk data lend themselves easily to a full quantitative analysis. For example, legal risk defies a precise quantitative analysis much more than, say, damage to physical assets. The analytic methods discussed cover a broad range of issues which may eventually enter in the development of an advanced measurement approach, AMA in the language of Basel II. Moreover, in the case of market and credit risk, we have witnessed a flurry of scientific activity around the various regulatory guidelines. Examples include the work on an axiomatic approach to risk measures and the development of advanced rating models for credit risk. This feedback from practice to theory can also be expected in the area of operational risk. Our paper shows some potential areas of future research. Under the AMA approach, banks will have to integrate internal data with relevant external loss data, account for stress scenarios, and include in the modelling process factors which reflect the business environment and the internal control system; see EBK (2005). Moreover, the resulting risk capital must correspond to a 99.9%-quantile (VaR) of the aggregated loss data over the period of a year. Concerning correlation, no specific rules are given (for instance within EBK (2005)) beyond the statement that explicit and implicit correlation assumptions between operational loss events as well as loss random variables used have to be plausible and need to be well founded.

In Section 2, we first present some more advanced techniques from the realm of extreme value theory (EVT). EVT is considered as a useful set of tools for analyzing rare events; several of the operational risk classes exhibit properties which in natural way call for an EVT analysis. To the ongoing discussion on the use of EVT, we add some techniques which could address non-stationarity in (some of) the underlying data.

In Section 3 we turn to the issue of dependence modelling. In a first instance, we assume no dependence information is given and, using the operational risk data introduced in Section 2, work out the so-called worst-VaR case for the aggregate data.

In Section 4, we then turn to the problem of modelling the interdependence between various operational risk processes. Here, several approaches are possible. We concentrate on one approach showing how copula-based techniques can be used to model dependent loss processes which are of the compound Poisson type. As already stated above, there is so far no agreement on how to model correlation. The methodology we offer is sufficiently general and contains many of the approaches already found in the literature on operational risk as special cases. This section puts some of these developments in a more structured context and indicates how future research on this important topic may develop further.

2 Advanced EVT Models

2.1 Why EVT?

The key attraction of EVT is that it offers a set of ready-made approaches to a challenging problem of quantitative (AMA) operational risk analysis, that is, how can risks that are both extreme and rare be modelled appropriately? Applying classical EVT to operational loss data however raises some difficult issues. The obstacles are not really due to a technical justification of EVT, but more to the nature of the data. As explained in Embrechts et al. (2003) and Embrechts et al. (2004), whereas EVT is the natural set of statistical techniques for estimating high quantiles of a loss distribution, this can be done with sufficient accuracy only when the data satisfy specific conditions; we further need sufficient data to calibrate the models. Embrechts et al. (2003) contains a simulation study indicating the sample size needed in order to reliably estimate certain high quantiles, and this under ideal (so called iid = independent and identically distributed) data structure assumptions. From the above two papers we can definitely infer that, though EVT is a highly useful tool for high-quantile (99.9%) estimation, the present data availability and data structure of operational risk losses make a straightforward EVT application somewhat questionable. Nevertheless, for specific subclasses where quantitative data can be reliably gathered, EVT offers a useful tool. An ever returning issue is the level at which a threshold has to be set beyond which the EVT asymptotics (in the GPD-POT form, say) can be fitted. There is no easy ready-made solution to this question. The issues raised in Diebold et al. (2001) still stand and some of them were raised before in Embrechts et al. (1997); see especially Figures 4.1.13, 6.4.11 and 5.5.4. The mathematical reasons why optimal threshold selection is very difficult indeed can best be appreciated by Example 4.1.12 in Embrechts et al. (1997): one needs second order information on the underlying (unknown) model. As soon as larger public databases on operational risk become available, hopefully more can be said on this issue. By now, numerous papers have been written on optimal threshold selection; see for instance Beirlant et al. (2004). We shall not address this issue further in this paper.

The current discussion on EVT applications to the AMA modelling of operational risk data will no doubt have an influence on the future research agenda in that field. Besides the threshold problem, other issues to be discussed already include risk capital point and interval estimation at (very) high quantile levels, the comparison of different estimation procedures (Hill, POT, DEdH etc.), the use of robust and Bayesian approaches to EVT, and EVT for non-stationary data. In this paper, we concentrate on the latter.

Consider Figure 1 taken from Embrechts et al. (2004). The data reflect a loss database of a bank for three different types of losses and span a 10-year period. The original data were transformed in such a way as to safeguard anonymity; the main characteristics however have been kept. This transformation unfortunately takes away the possibility to discuss the underlying practical issues at any greater length. For our paper we therefore only discuss the resulting statistical modelling issues. Also note that the number of losses is fairly small.

Besides the apparent existence of extremes (hence EVT matters), the data seem to increase in frequency over time, with a fairly radical change around 1998 (hence non-stationarity may be an issue). One obvious reason for this apparent change-in-frequency could be that quantification of operational risk only became an issue in the late nineties. This is referred to as *reporting bias*. Such structural changes may also be due to an internal change (endogenous

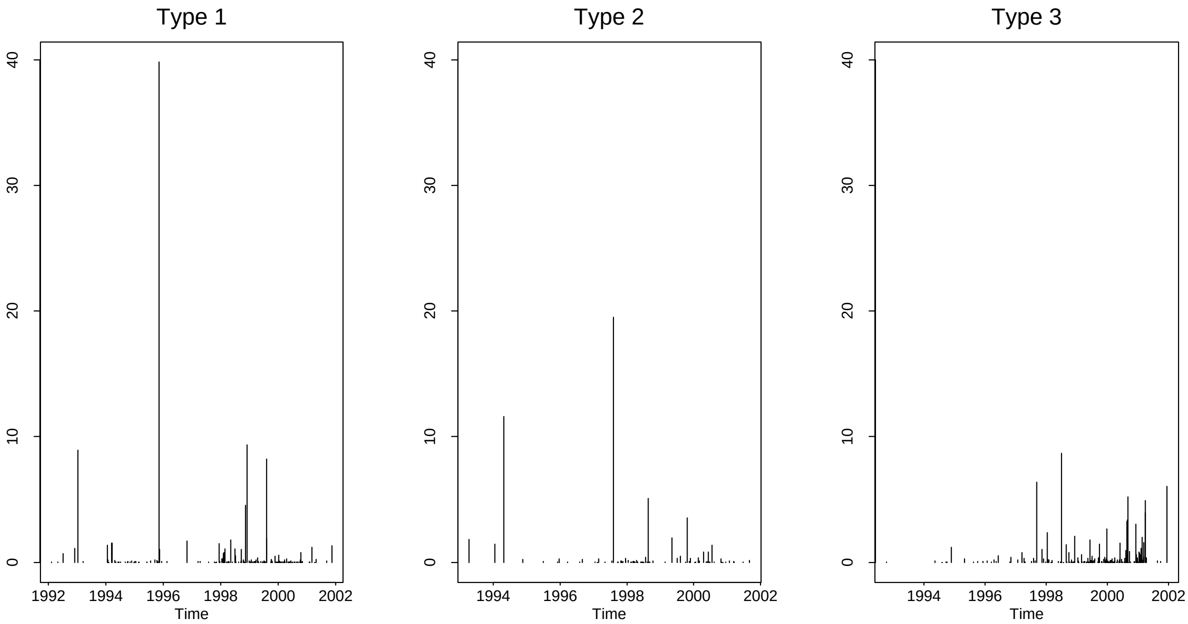

Figure 1: Operational risk losses. From left to right: Type 1 ($n = 162$), Type 2 ($n = 80$), Type 3 ($n = 175$).

effects, management action, M&A) or changes in the economic/political/regulatory environment in which the company operates (exogenous effects). As far as we are aware, no detailed studies reflecting non-stationarity in the loss frequency exist. From the calculation of a global risk measure for operational risk, we expect however that the frequency issue is secondary. Especially for the bigger losses, a clean time stamp may not be readily available, also banks use clumping of smaller claims at the end of certain time periods. Nevertheless, we find it useful to show how EVT has been extended taking more general loss frequency processes into account.

We adapt classical EVT to take both non-stationarity and covariate modelling (different types of losses) into account. Chavez-Demoulin (1999), Chavez-Demoulin and Davison (2005) contain the relevant methodology. In the next subsection, we first review the Peaks over Threshold (POT) method and the main operational risk measures to be analysed. In Subsection 2.3, the adapted classical POT method, taking non-stationarity and covariate modelling into account, is applied to the operational risk loss data from Figure 1.

2.2 The basic EVT methodology

Over the recent years, EVT has been recognized as a useful set of probabilistic and statistical tools for the modelling of rare events and its impact on insurance, finance and quantitative risk management is well recognized. Numerous publications have exemplified this point. Embrechts et al. (1997) detail the mathematical theory with a number of applications to insurance and finance. The edited volume Embrechts (2000) contains an early summary of EVT applications to risk management, whereas McNeil et al. (2005) contains a concise discussion with quantitative risk management applications in mind. Reiss and Thomas (2001), Falk et al. (2004), Coles (2001) and Beirlant et al. (2004) are very readable introductions to EVT in general. Numerous papers have looked at EVT applied to operational risk; see for instance Moscadelli (2004), Cruz (2004) and the references therein.

Below, we give a very brief introduction to EVT and in particular to the peaks over threshold (POT) method for high-quantile estimation. A more detailed account is to be found in the list of references; for our purpose, i.e. the modelling of non-stationarity, Chavez-Demoulin and Davison (2005) and Chavez-Demoulin and Embrechts (2004) contain relevant methodological details.

From the latter paper, we borrow the basic notation (see also Figure 2):

- ground-up losses are denoted by $Z_1, Z_2, \ldots, Z_q$;
- u is a typically high threshold, and
- $W_1, \ldots, W_n$ are the excess losses from $Z_1, \ldots, Z_q$ above u, i.e. $W_j = Z_i - u$ for some $j = 1, \ldots, n$ and $i = 1, \ldots, q$, where $Z_i > u$.

Note that u is a pivotal parameter to be set by the modeller so that the excesses above u, $W_1, \ldots, W_n$, satisfy the required properties from the POT method; see Leadbetter (1991) for the basic theory. The choice of an appropriate u poses several difficult issues in the modelling of operational risk; see the various discussions at a meeting organized by the Federal Reserve Bank of Boston, Implementing an AMA for Operational Risk, Boston, May 18–20, 2005 (www.bos.frb.org/bankinfo/conevent/oprisk2005) and the brief discussion in Section 1. For iid losses, the conditional excesses $W_1, \ldots, W_n$, asymptotically for u large, follow a so-called

Generalized Pareto Distribution (GPD):

$$G_{\kappa,\sigma}(w) = \begin{cases} 1-(1+\kappa w/\sigma)_+^{-1/\kappa}\,, & \kappa \neq 0, \\ 1-\exp(-w/\sigma)\,, & \kappa = 0\,, \end{cases} \tag{1}$$

where $(x)_+ = x$ if $x > 0$ and 0 otherwise. The precise meaning of the asymptotics is explained in Embrechts et al. (1997), Theorem 3.4.13. Operational loss data seem to support $\kappa > 0$ which corresponds to ground-up losses $Z_1, \dots, Z_q$ following a Pareto-type distribution with power tail with index $1/\kappa$, i.e. $P(W_i > w) = w^{-1/\kappa}h(w)$ for some slowly varying function h, i.e. h satisfies

$$\lim_{t\to\infty} \frac{h(tw)}{h(t)} = 1, \quad w > 0. \tag{2}$$

For instance, in a detailed study of all the losses reported to the Basel Committee during the third Quantitative Impact Study (QIS), Moscadelli (2004) finds typical Pareto-type behavior across most of the risk types, even some cases with $\kappa > 1$, i.e. infinite mean models.

From Leadbetter (1991) it also follows that for u high enough, the exceedance points of $Z_1, \dots, Z_q$ of the threshold u follow (approximately) a homogeneous Poisson process with intensity $\lambda > 0$. Based on this, an approximate log-likelihood function $l(\lambda, \sigma, \kappa)$ can be derived; see Chavez-Demoulin and Embrechts (2004) for details. In many applications, including the modelling of operational risk, it may be useful to allow the parameters λ, σ, κ in the POT method to be dependent on time and explanatory variables allowing for non-stationarity. In the next section (where we apply the POT method to the data in Figure 1), we will take for $\lambda = \lambda(t)$ a specific function of time which models the apparent increase in loss intensity in Figure 1. We moreover will differentiate between the different loss types and adjust the severity loss parameters κ and σ accordingly.

Basel II requires banks using the AMA to measure risk using a one-year 99.9 % Value-at-Risk. In Section 3 we will discuss some of the consequences coming from this requirement. For the moment it suffices to accept that the high-quantile level of 99.9 % opens the door for EVT methodology. Given a loss distribution F, we denote $\mathrm{VaR}_\alpha = F^{-1}(\alpha)$, where F^{-1} can be replaced by the generalized inverse $F^{\leftarrow}$ when necessary; see Embrechts et al. (1997), p. 130. In cases where the POT method can be applied, for given u, this measure can be estimated as follows:

$$\widehat{\mathrm{VaR}}_\alpha = u + \frac{\hat\sigma}{\hat\kappa}\left\{\left(\frac{1-\alpha}{\hat\lambda}\right)^{-\hat\kappa} - 1\right\}. \tag{3}$$

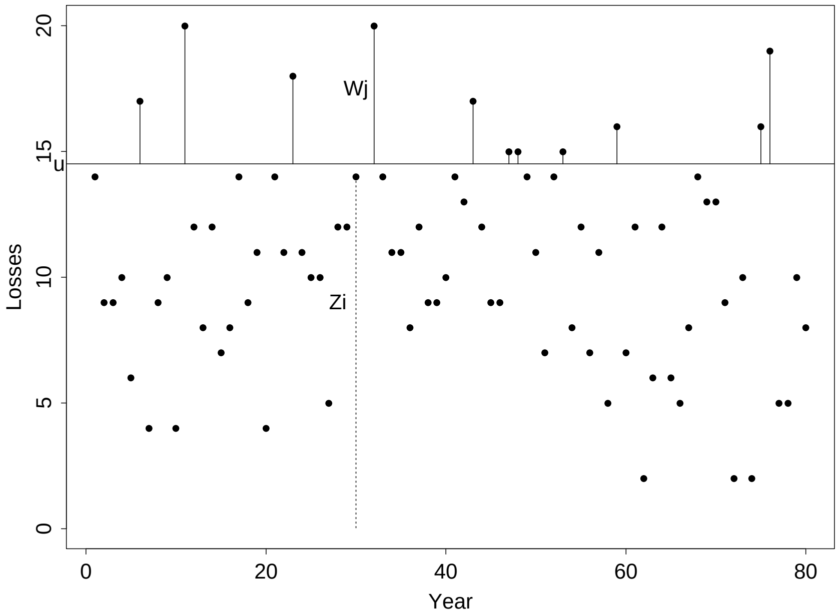

Figure 2: The point process of exceedances (POT).

Here $\hat{\lambda}, \hat{\kappa}, \hat{\sigma}$ are the maximum likelihood estimators of λ, κ and σ. Interval estimates can be obtained by the delta method or by the profile likelihood approach and has been programmed for instance into the freeware EVIS by Alexander McNeil, available under www.math.ethz.ch/~mcneil; see McNeil et al. (2005) for details.

2.3 POT analysis of the operational loss data

Consider the operational risk data of Figure 1 pooled across the three risk types. The main features of the pooled data hence are risk type, extremes and indication of non-stationarity in the loss-frequency. Consequently, any EVT analysis of the pooled data should at least take the risk type τ as well as the non-stationarity (switch around 1998, say) into account. Using the advanced POT modelling, including non-stationarity and covariates, the data pooling has the advantage to allow for testing interaction between explanatory variables: is there for instance an interaction between type of loss and change in frequency, say? In line with Chavez-Demoulin and Embrechts (2004), we fix a threshold $u = 0.4$. The latter paper also contains a sensitivity analysis of the results with respect to this choice of threshold u, though again, the same fundamental issues exist concerning an "optimal" choice of u. A result from such an analysis is that for these data, small variations in the value of the threshold have no significant impact.

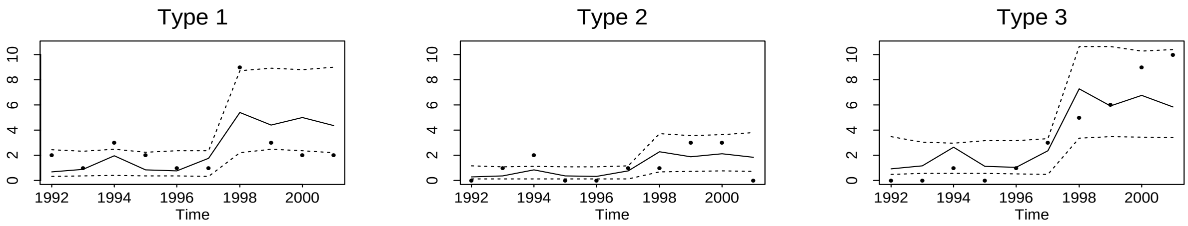

Figure 3: Operational risk losses. From left to right: Estimated Poisson intensity $\hat{\lambda}$ and 95% confidence intervals for data of loss type 1, 2, 3. The points are the yearly numbers of exceedances over $u = 0.4$.

Following the non-parametric methodology summarized in the above paper, we fit different models for λ, κ and σ allowing for:

- functional dependence on *time* $g(t)$, where t refers to the *year* over the period of study;
- dependence on τ, where τ defines the *type* of loss data through an indicator $I_\tau = 1$, if the type equals τ and 0 otherwise, with $\tau = 1, 2, 3$, and
- *discontinuity* modelling through an indicator $I_{(t>t_c)}$ where $t_c = 1998$ is the year of possible change point in the frequency and

$$I_{(t>t_c)} = \begin{cases} 1, & \text{if} \quad t > t_c, \\ 0, & \text{if} \quad t \leq t_c. \end{cases}$$

Of course a more formal test on the existence and value of t_c can be incorporated. We apply different possible models to each parameter λ, κ and σ. Using specific tests (based on the likelihood ratio statistics), we compare the resulting models and select the most significant one.

The selected model for the Poisson intensity $\lambda(t, \tau)$ turns out to be

$$\log \hat{\lambda}(t, \tau) = \hat{\gamma}_\tau I_\tau + \hat{\beta} I_{(t>t_c)} + \hat{g}(t). \tag{4}$$

Inclusion of the first component $\hat{\gamma}_\tau I_\tau$ on the right hand side indicates that the type of loss τ is important to model the Poisson intensity; that is the number of exceedances over the threshold differs significantly for each type of loss 1, 2 or 3. The selected model also contains the discontinuity indicator $I_{(t>t_c)}$ as a test based on the hypothesis that the simplest model "$\beta = 0$ suffices" is rejected at a 5% level. We find $\hat{\beta} = 0.47(0.069)$ and the intensity is rather

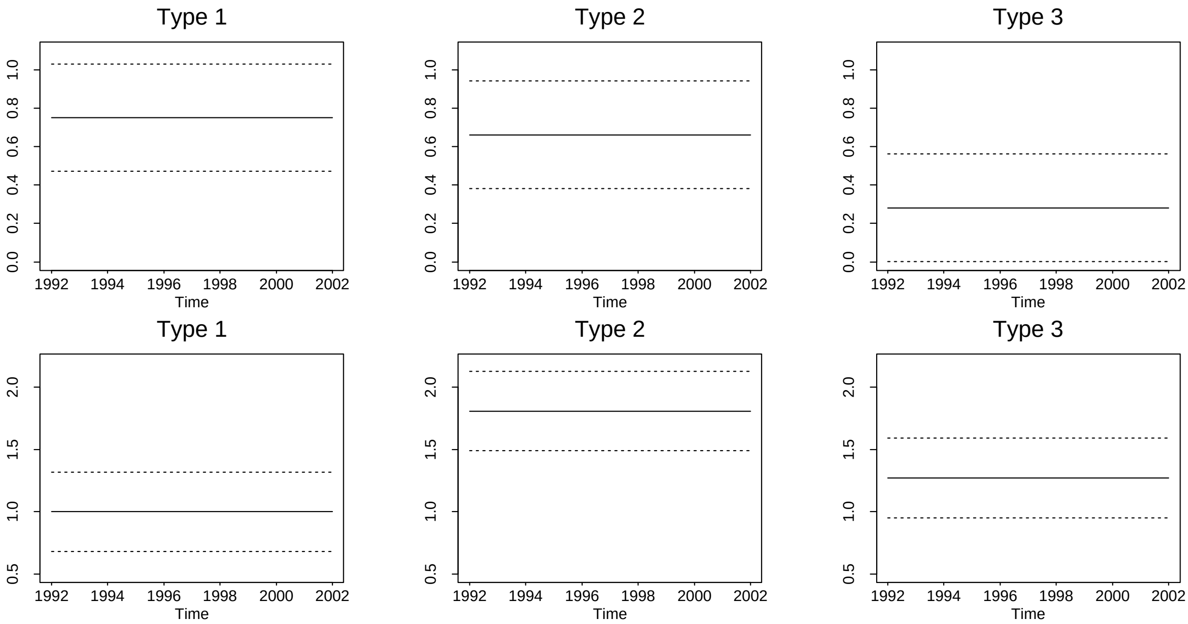

Figure 4: Estimated GPD parameters for the operational risk losses from Figure 1: upper $\hat{\kappa}$, lower $\hat{\sigma}$ and 95% confidence intervals for different loss types.

different in mean before and after 1998. Finally, it is clear that the loss intensity parameter λ is dependent on time (year). This dependence is modelled through the estimated function $\hat{g}(t)$. For the reader interested in fitting details, we use a smoothing spline with 3 degrees of freedom selected by AIC (see Chavez-Demoulin and Embrechts (2004)); see also Green and Silverman (1994) for further support on the use of cubic splines. Figure 3 represents the resulting estimated intensity $\hat{\lambda}$ for each type of losses and its 95% confidence interval based on bootstrap resampling schemes (details in Chavez-Demoulin and Davison (2005)). The resulting curves seem to capture the behaviour of the number of exceedances (points of the graphs) for each type rather well. The method would also allow to detect any seasonality or cyclic patterns which may exist; see Brown and Wang (2005). Similarly, we fit several models for the GPD parameters $\kappa = \kappa(t, \tau)$ and $\sigma = \sigma(t, \tau)$ modelling the losssize through (1) and compare them. For both κ and σ, the model selected depends only on the type τ of the losses but not on time t. Their estimates $\hat{\kappa}(\tau)$ and $\hat{\sigma}(\tau)$ and 95% confidence intervals are given in Figure 4. Point estimates for the shape parameter κ (upper panels) are $\kappa_1 = 0.7504$, $\kappa_2 = 0.6607$ and $\kappa_3 = 0.2815$; this suggests a loss distribution for type 3 with a less heavy tail than for types 1 and 2. Tests based on likelihood ratio statistics have shown that the effect due to the switch in 1998 is not retained in the models for κ and σ, i.e. the loss size distributions do not switch around 1998. Finally, note that, as the GPD parameters κ and σ are much harder to estimate than λ, the lack of sufficient data makes the detection of any trend and/or periodic components difficult. Also for this reason, the resulting 95 % confidence intervals are wide.

To assess the model goodness-of-fit for the GPD parameters, a possible diagnostics can be based on the result that, when the model is correct, the residuals

$$R_j = \hat{\kappa}^{-1} \log\{1 + \hat{\kappa} W_j / \hat{\sigma}\}, \qquad j = 1, \ldots, n, \tag{5}$$

are approximately independent, unit exponential variables. Figure 5 gives an exponential quantile-quantile plot for the residuals using the estimates $\hat{\kappa}(\tau)$ and $\hat{\sigma}(\tau)$ for the three types of loss data superimposed. This plot suggests that our model is reasonable.

The potential importance of using models including covariates (representing type) instead of pooling the data and finding unique overall estimated values of λ, κ, σ is clearly highlighted here. In a certain sense, the use of our adapted model allows to exploit all the information available on the data, a feature which is becoming more and more crucial, particularly in the context of operational and credit risk. Other applications may be found at the level of a regulator where pooling across different banks may be envisaged or for comparing and contrasting internal versus external loss data. Using the estimated parameters $\hat{\lambda}, \hat{\kappa}, \hat{\sigma}$ it is possible to estimate VaR (see (3)) or related risk capital measures; for this to be done accurately much larger data bases must become available. The data displayed in Figure 1 are insufficient for such an estimation procedure at the 99.9 % confidence level, leading to very wide confidence regions for the resulting risk capital.

3 Aggregating (Operational) Risk Measures

3.1 The risk aggregation problem; an example

The risk aggregation issue for operational risk in the Advanced Measurement Approach within the Basel II framework typically, though not exclusively, starts with a number d (7 risk types, 8 business lines, 56 classes) of loss random variables $L_1, \ldots, L_d$ giving the total loss amount for a particular type/line/class for the next accounting year, say. By the nature of operational risk data (see Section 2), these random variables are often of the type $L_k = \sum_{i=1}^{N_k} X_i(k)$,

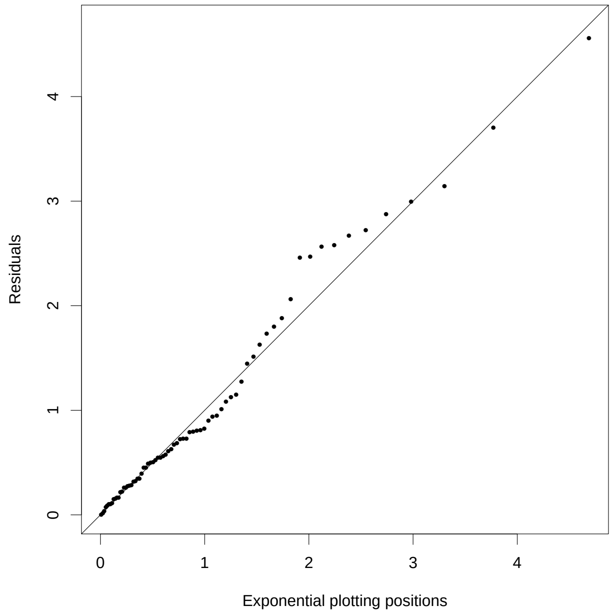

Figure 5: QQ-plot of fitted residuals (5) against exponential plotting positions.

$k = 1, \ldots, d$ where N_k is a frequency random variable assumed to be independent of the iid severity random variables $X_i(k)$; the rvs L_k are referred to as of the *compound type*. If N_k has a Poisson distribution, then L_k is called a *compound Poisson* random variable. For general compound rvs and heavy-tailed severity distributions, it is known that $\mathsf{P}(L_k > x)$ inherits the tail-properties of $\mathsf{P}(X_i(k) > x)$, in particular, if $\mathsf{P}(X_i(k) > x) = x^{-1/\kappa_k} h_k(x)$ for some slowly varying function h_k and $\mathrm{E}(N_k) < \infty$, then $\lim_{x\to\infty} \mathsf{P}(L_k > x)/\mathsf{P}(X_i(k) > x) = \mathrm{E}(N_k)$. For the exact conditions on N_k, see Embrechts et al. (1997), Theorem A 3.20. Although the precise modelling of N_k in the case of Figure 1 needs more data, it seems reasonable to assume, as a first approximation, that the tail behavior of $\mathsf{P}(L_k > x)$ is Pareto-type with index κ_k, $k = 1, 2, 3$ as given in Section 2.3. To highlight the problem at hand, and for notational convenience, we will assume more precisely that

$$\mathsf{P}(L_k > x) = (1 + x)^{-1/\kappa_k}, \quad \kappa \geq 0, \quad k = 1, 2, 3, \tag{6}$$

and denote the corresponding VaR measures by $\mathrm{VaR}_\alpha(k)$ where we are typically interested in the α-range $(0.99, 0.9999)$, say. Note that we left out the (asymptotic) frequency correction $\mathrm{E}(N_k)$; on this issue, more work is needed.

In this particular case, $d = 3$, the total loss to be modelled is $L = \sum_{k=1}^{3} L_k$; this random variable in general may be very complex as it typically contains components with rather different frequency as well as severity characteristics. Moreover, the interdependence between

α	$\mathrm{VaR}_\alpha(L_1)$	$\mathrm{VaR}_\alpha(L_2)$	$\mathrm{VaR}_\alpha(L_3)$	I	C	DB	SB
0.9	4.6	3.6	0.9	8.8	9.1	18.0	18.5
0.99	30.7	19.9	2.7	43.6	53.3	90.5	93.4
0.999	177.3	95.0	6.0	165.3	278.3	453.1	461.5
0.9999	1002.7	438.3	12.4	299.9	1453.4	2303.5	2327.7

Table 1: Bounds on Value-at-Risk: comonotonicity (C), independence (I), dual bound (DB) and standard bound (SB).

the various L_k's is largely unknown. In Section 4, we will return to this important issue. Under the Basel II AMA guidelines for operational risk, a capital charge can be calculated as $\sum_{k=1}^{3} \mathrm{VaR}_{99.9\%}(L_k)$. However, because of the possible non-coherence of VaR, it is not clear whether indeed subadditivity holds in this case, i.e. whether

$$\mathrm{VaR}_{99.9\%}(L) = \mathrm{VaR}_{99.9\%}\left(\sum_{k=1}^{3} L_k\right) \leq \sum_{k=1}^{3} \mathrm{VaR}_{99.9\%}(L_k). \tag{7}$$

Indeed, the typical examples where the inequality ($\leq$) may reverse ($>$) occur when the distribution functions of the L_k's are very skewed, when the rvs L_k have a very special dependence or when the underlying distribution functions are (very) heavy-tailed; see McNeil et al. (2005), Chapter 6 for details and references to the underlying examples. For our purposes, i.e. for the quantitative modelling of operational risk, all three potential non-coherence preconditions are relevant. Hence the important question then becomes: by how much can (7) be violated? In particular, given the marginal loss distributions (6), what is the maximal value of the risk capital $\mathrm{VaR}_{99.9\%}\left(\sum_{k=1}^{3} L_k\right)$ under all possible dependence assumptions for the loss vector (L_1, L_2, L_3). Though this question is in general difficult to answer, there are several numerical solutions yielding bounds to this quantity. A detailed discussion on the background of such so-called Fréchet-type problems is to be found in Puccetti (2005) and in McNeil et al. (2005). For our purposes, relevant is Embrechts and Puccetti (2006b) and in particular Embrechts and Puccetti (2006a). The latter paper computes numerically upper bounds for $\mathrm{VaR}_\alpha\left(\sum_{i=1}^{d} X_i\right)$ for d one-period risks $X_1, \dots, X_d$ with possibly different dfs F_{X_i}, $i = 1, \dots, d$. This is computationally non-trivial; in that paper it is shown how a so-called dual bound (DB in Table 1) improves on the easier-to-calculate standard bound (SB). For this terminology from the theory of Fréchet problems and Mass Transportation Theory, we refer to Embrechts and Puccetti

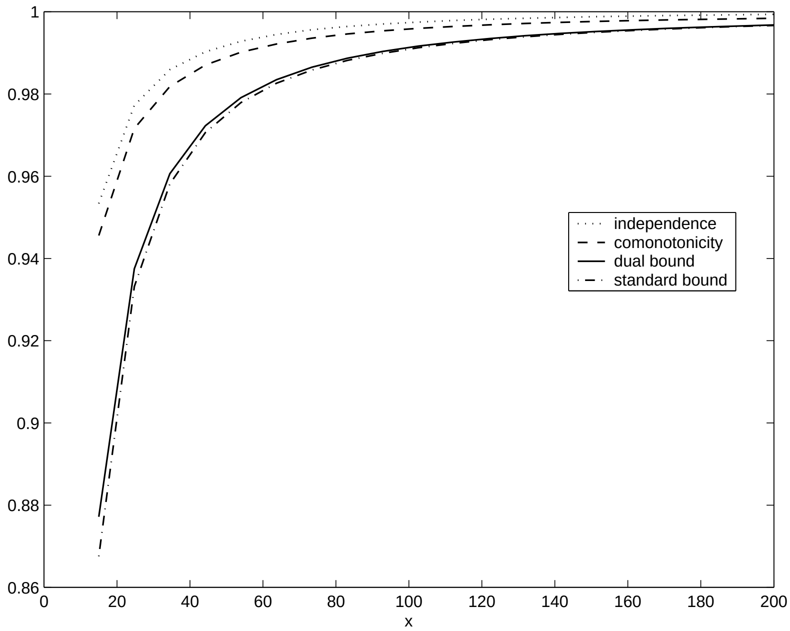

Figure 6: Bounds on the df $\mathsf{P}(L_1+L_2+L_3 \leq x)$ for the Pareto loss dfs of Section 2.3, i.e. (6).

(2006a) as well as Rachev and Rüschendorf (1998a) and Rachev and Rüschendorf (1998b).

In Table 1 we have summarized the results from the above optimization problem in the case of the assumption (6) for the loss rvs L_1, L_2, L_3 from Figure 1, i.e. with κ_1, κ_2 and κ_3 as given in Section 2.3. We report the standard and (improved) dual bounds computed using the results from Embrechts and Puccetti (2006a) and compare and contrast these bounds with the exact values under the assumptions of independence and comonotonicity. Recall that L_1, L_2, L_3 are comonotone if there exist a rv Z and increasing functions f_1, f_2, f_3 so that $L_k = f_k(Z)$, $k = 1, 2, 3$. Under this strong assumption of (so-called perfect) dependence, Value-at-Risk is additive, i.e. $\mathrm{VaR}_\alpha\left(\sum_{k=1}^3 L_k\right) = \sum_{k=1}^3 \mathrm{VaR}_\alpha(L_k)$; see McNeil et al. (2005), Proposition 6.15. So for example if $\alpha = 0.999$ and (6) is assumed, one immediately finds that $\mathrm{VaR}_\alpha(L_1) = 177.3$, $\mathrm{VaR}_\alpha(L_2) = 95.0$ and $\mathrm{VaR}_\alpha(L_3) = 6.0$, so that under the assumption of comonotonicity, $\mathrm{VaR}_{99.9\%}\left(\sum_{k=1}^3 L_k\right) = 278.3$, as reported in Table 1. For L_1, L_2, L_3 independent, one finds the value 165.3, whereas the dual bound on $\mathrm{VaR}_{99.9\%}(L)$ equals 453.1. Though the construction of sharp bounds for $d \geq 3$ is still an open problem, using copula techniques as explained in Embrechts et al. (2005), one can construct models for (L_1, L_2, L_3) with VaR-values in the interval $(278.3, 453.1)$. Note that VaR-values in the latter interval always correspond to a non-coherence (i.e. non-subadditive) situation for Value-at-Risk. Figure 3.1 contains a graphical presentation of the results in terms of the df $\mathsf{P}(L_1 + L_2 + L_3 \leq x)$.

3.2 Discussion

As already stated above, the loss characteristics of operational loss data, as summarized by heavy-tailedness, skewness and unknown interdependence between the various loss rvs, imply that the Value-at-Risk measure for risk capital may not be subadditive. Due to a lack of publicly available data, it is not yet clear to what extend correlation issues can be taken into account which may lead to a reduction of the calculated risk capital based on $\sum_{k=1}^{d} \mathrm{VaR}_{99.9\,\%}(L_k)$. The results from Embrechts and Puccetti (2006a) as exemplified in Section 3.1, yield an upper bound for the worst case. The tools of Section 4 may help in understanding the modelling of the dependence between the compound rvs $L_1, \ldots, L_d$. By nature of the (loss) data, the loss dfs are typically skewed to the right. This leaves the Pareto-type (power) behavior of $\mathsf{P}(L_k > x)$, $k = 1, \ldots, d$ as a distinct possibility. Support for this assumption is obtained from Moscadelli (2004) and in part from de Fontnouvelle (2005). Again, these preliminary statistical analyses on summarised banking industry data cannot be considered as a proof of power-like behavior, the several results we have seen however contain a strong indication in that direction. It will be of crucial importance to investigate this more deeply in the (near) future; especially the issue of infinite mean GPD models for most of the Basel II business lines, as reported in Moscadelli (2004), calls for special attention. The relatively little data underlying Figure 1, though obviously heavy-tailed, does not allow for a statistically conclusive answer to this issue. The reader interested in consequences of extreme heavy-tailedness of portfolio losses is advised to consult Embrechts et al. (1997), especially Chapter 1, Section 8.2. and 8.3, and Asmussen (2000); look for the “one loss causes ruin” problem on p. 264. For strong support on and consequences of the power-tail behavior in finance, see for instance Mandelbrot (1997) and Rachev et al. (2005).

4 Dependent Risk Processes

4.1 The point process approach

Apart from handling non-stationarity and extremes in operational loss data, the understanding of diversification effects in operational risk modelling is of key importance, especially in the light of the discussions of the previous section. For each of d risk types to be modelled, one may obtain operational loss series; for the purpose of this section assume that we are able to model them appropriately. It is however intuitively clear that risk events may be related across different classes. Consider for example effects with a broad impact, such as mainframe or electricity failure, weather catastrophes, major economic events or terrorist attacks like September 11. On such severe occasions, several business lines will typically be affected and cause simultaneous losses of different risk types.

In this section, we present two methods for modelling dependent loss processes based on Pfeifer and Nešlehová (2004). A key point here is to view loss processes in an equivalent, yet mathematically more tractable way, namely as *point processes*. This approach may appear less appealing at first sight because of its rather complicated theoretical background. This is however more then compensated for by the clear advantages it has when it comes to more advanced modelling. In the context of EVT for instance, the point process characterization not only unifies several well-known models such as block maxima or threshold exceedances but also provides a more natural formulation of non-stationarity; see McNeil et al. (2005), Coles (2001) and especially Resnick (1987). The techniques presented in Section 2 very much rely on point process methodology. Point process theory also forms the basis for the intensity based approach to credit risk; see Bielecki and Rutkowski (2002). A detailed discussion of the use of point process methodology for the modelling of multivariate extremes (multivariate GPDs and threshold models) with the modelling of so-called high risk scenarios in mind, is Balkema and Embrechts (2006). In this section, we show that also the issue of dependence can be tackled in a very general, though elegant way when using this methodology. We also show how recent models proposed for describing dependence within operational loss data can be viewed as special cases.

To lessen the theoretical difficulties, we devote this subsection to an informal introduction to the basics of the theory of point processes in the context of operational risk. For information beyond this brief introduction, we refer to Chapter 5 in Embrechts et al. (1997), Reiss (1993), Kingman (1993) or the comprehensive monograph by Daley and Vere-Jones (2003).

The key ingredients of loss data in operational, credit and underwriting risk, for instance, are the occurrence of the event and the loss size/severity. We first concentrate on the occurrences, i.e. frequency (see Subsection 4.4 for the severities). Loss occurrences will typically follow a Poisson counting process; the aim of the discussion below is to show that an alternative representation as a point process is possible, which more naturally allows for dependence.

Suppose that a loss event happens at a random time T in some period under study $[0,\Delta]$, say. In our case, Δ will typically be one (year). For every set $A \subset [0,\Delta]$, we can evaluate the easiest point process I_T:

$$I_T(A) = \begin{cases} 1, & \text{if } T \in A, \\ 0, & \text{otherwise,} \end{cases}$$

also referred to as an elementary *random measure*. Next, let $T_1, \dots T_n$ be n random loss events, then the point process ξ_n given by

$$\xi_n(A) := \sum_{i=1}^{n} I_{T_i}(A) \tag{8}$$

counts the number of losses in the observation period $A \subset [0,\Delta]$. There are several ways in which we can generalize (8) in order to come closer to situations we may encounter in reality. First, we can make n random, N say, which leads to a random number of losses in $[0,\Delta]$. In addition, the T_i's can be multivariate, $\boldsymbol{T}_i$ d-dimensional, say. The latter corresponds to occurrences of d loss types (all caused by one effect for instance). This leads to the general point process

$$\xi_N := \sum_{i=1}^{N} I_{\boldsymbol{T}_i}. \tag{9}$$

Recall that all components of $\boldsymbol{T}_i$ are assumed to lie in $[0,\Delta]$, i.e. $\xi_N([0,\Delta]^d) = N$. As a special case, consider $d = 1$ and N Poisson with parameter $\lambda\Delta$ and independent of the T_i's, which themselves are assumed mutually independent and uniformly distributed on $[0,\Delta]$. If $A = [0,t]$ for some $0 \le t \le \Delta$, then one can verify that

$$\big\{N(t) := \xi_N\big([0,t]\big) : t \in [0,\Delta]\big\}$$

is the well known homogeneous Poisson counting process with rate (intensity) $\lambda > 0$, restricted to $[0,\Delta]$. Recall that in this case

$$\mathrm{E}\big(N(t)\big) = \mathrm{E}(N)\mathsf{P}[T_i \le t] = \lambda\Delta\frac{t}{\Delta} = \lambda t.$$

Note that, in contrast to the classical construction of $\{N(t) : t \geq 0\}$ as a renewal process, the sequence of the loss occurrence times T_i is not necessarily ascending. The restriction to the finite time period $[0, \Delta]$, which is not needed in the traditional counting process approach, can also be overcome in the point process world; we come back to this issue in the discussion below.

The advantage of the point process modelling now becomes apparent as it naturally leads to further generalizations. The time points can still occur randomly in time, but with a time variable intensity. Moreover, the loss occurrences can be d-dimensional like in (9), or replaced by $(\boldsymbol{T}_i, \boldsymbol{X}_i)$ where the $\boldsymbol{X}_i$'s denote the corresponding severities (see Subsection 4.4). Note however that to this point, we assume the total number of losses to be the same for each component. A construction method which relaxes this will be the subject of Subsection 4.3. If the common counting variable N has a Poisson distribution and is independent of the iid loss occurrences, which follow some unspecified distribution F, then (9) is a (finite) *Poisson point process*, which we from now on denote by ξ. In that case $\xi(A)$ is an ordinary Poisson random variable with parameter $\mathrm{E}\,\xi(A) = \mathrm{E}(N)F(A)$. As a function of A, $\mathrm{E}\,\xi(\cdot)$ is referred to as the *intensity measure* of ξ. Whenever this measure has a density then this is called the *intensity* of the point process. Moreover, if $A_1, \ldots, A_n$ are mutually disjoint time intervals, the numbers of occurrences within those intervals, $\xi(A_1), \ldots, \xi(A_n)$, are independent.

From now on assume that the process of loss occurrences is a Poisson point process of the form (9). Below, we list three properties of Poisson point processes which are key for modelling dependence; for proofs and further details, we refer to the literature above.

Let $\xi = \sum_{i=1}^{N} I_{\boldsymbol{T}_i}$ be a finite Poisson point process with d-dimensional event points $\boldsymbol{T}_i = (T_i(1), \ldots, T_i(d))$. For example, for $d = 2$, $T_i(1)$ and $T_i(2)$ can denote occurrence time points of losses due to internal and external fraud in the same business line. Each of the *projections* or, marginal processes,

$$\xi(k) = \sum_{i=1}^{N} I_{T_i(k)}, \quad k = 1, \ldots, d, \tag{10}$$

is then a one-dimensional Poisson point process, i.e. a process describing internal and external fraud losses, respectively. The intensity measure $\mathrm{E}\,\xi(k)(\cdot)$ of the marginal processes is given by $\mathrm{E}(N)F_k(\cdot)$, where F_k denotes the k-th margin of the joint distribution F of the T_i. Figure 7 (left) shows a two-dimensional homogeneous Poisson point process with intensity 20. The one-dimensional projections are displayed on the axes as well as in Figure 7 (right).

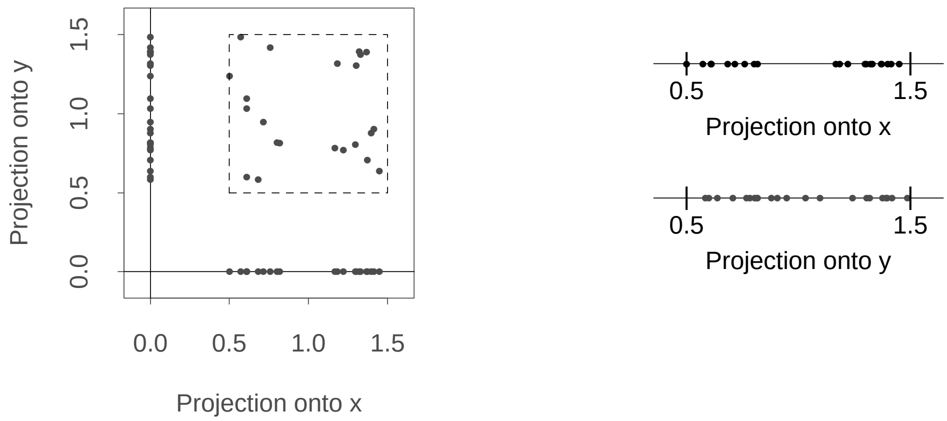

Figure 7: Projections of a two dimensional homogeneous Poisson point process on $[0.5, 1.5] \times [0.5, 1.5]$.

Conversely, if $\xi(k) = \sum_{i=1}^{N} I_{T_i(k)}$, $k = 1, \ldots, d$, are one-dimensional Poisson point processes, then $\xi = \sum_{i=1}^{N} I_{\boldsymbol{T}_i}$ with $\boldsymbol{T}_i = (T_i(1), \ldots, T_i(d))$ is a d-dimensional Poisson point process with intensity measure $\mathrm{E}\,\xi(\cdot) = \mathrm{E}(N)F(\cdot)$ where F denotes the joint distribution of $\boldsymbol{T}_i$. This result, also called *embedding*, is of particular use for modelling dependent losses triggered by a common effect, as we will soon see.

Above, we considered only Poisson point processes on a finite period of time $[0, \Delta]$. It is however sometimes necessary to work on an infinite time horizon, such as e.g. $[0, \infty)$. To accomplish this, the definition of Poisson point processes can be extended, see e.g. Embrechts et al. (1997) or Reiss (1993). The resulting process is no longer given by the sum (9), but can be expressed as a sum of finite Poisson processes, a so-called *superposition*. Let ξ_1 and ξ_2 be independent Poisson point processes with (finite) intensity measures $\mathrm{E}\,\xi_1$ and $\mathrm{E}\,\xi_2$. Then the *superposition* of ξ_1 and ξ_2, i.e. the process $\xi = \xi_1 + \xi_2$, is again a Poisson point process with intensity measure $\mathrm{E}\,\xi = \mathrm{E}\,\xi_2 + \mathrm{E}\,\xi_1$.

Figure 8 shows a superposition of two homogeneous Poisson processes with different intensities defined on different time intervals. Another example would be a superposition of independent Poisson processes corresponding to different risk classes over the same time period (see Figure 13). Extending this result to a superposition of countably many independent Poisson processes yields a Poisson point process (in a wider sense) with an intensity measure that is

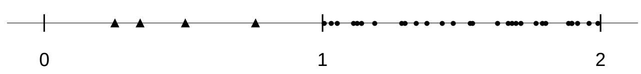

Figure 8: Superposition of a homogeneous Poisson process with intensity 5 over $[0,1]$ and a homogeneous Poisson process with intensity 20 over $[1,2]$.

not necessarily finite. For example, if ξ_k is a homogeneous Poisson point process with constant intensity $\lambda > 0$ (independent of k) on $[k-1,k)$ for a non-negative integer k, then the superposition $\xi = \sum_{k=1}^{\infty} \xi_k$ is a (locally) homogeneous Poisson point process on $[0,\infty)$. It moreover corresponds to the classical time-homogeneous Poisson counting process or renewal counting process with iid random interarrival times following an exponential distribution with expectation $1/\lambda$.

A final important technique is *thinning*, which splits a Poisson point process into two (or more) independent Poisson processes. It is accomplished by marking the event points with "1" or "0" using a random number generator and subsequent grouping of the event time points with identical marks. For instance, considering the point process of exceedances over a threshold u, we can mark by "1" those losses which exceed an even higher threshold $u + x$. Suppose $\xi = \sum_{i=1}^{N} I_{\boldsymbol{T}_i}$ is some (finite) Poisson point process and $\{\varepsilon_i\}$ a sequence of iid $\{0,1\}$-valued random variables with $\mathsf{P}[\varepsilon_i = 1] = p$. Then the thinnings of ξ are point processes given by

$$\xi_1 := \sum_{i=1}^{N} \varepsilon_i \cdot I_{\boldsymbol{T}_i} \quad \text{and} \quad \xi_2 := \sum_{i=1}^{N} (1-\varepsilon_i) \cdot I_{\boldsymbol{T}_i}. \tag{11}$$

The so-constructed processes ξ_1 and ξ_2 are independent Poisson point processes with intensities $\mathrm{E}\,\xi_1 = p\,\mathrm{E}\,\xi$ and $\mathrm{E}\,\xi_1 = (1-p)\,\mathrm{E}\,\xi$. Moreover, the original process arises as a superposition of the thinnings, $\xi = \xi_1 + \xi_2$.

As we will soon see, there are two kinds of dependence which play an important role for the Poisson point processes, $\xi_1 = \sum_{i=1}^{N_1} I_{T_i(1)}$ and $\xi_2 = \sum_{i=1}^{N_2} I_{T_i(2)}$, say:

- dependence between the events such as time occurrences of losses, e.g. between $T_i(1)$ and $T_i(2)$, and
- dependence between the number of events or event frequencies, e.g. between the counting (Poisson distributed) random variables N_1 and N_2.

Before presenting the models for dependent Poisson point processes, we first address these two issues.

4.2 Dependent counting variables

Modelling of multivariate distributions with given marginals can be accomplished in a particularly elegant way using *copulas*. This approach is based upon the well-known result of Sklar that any d-dimensional distribution function F with marginals $F_1, \ldots, F_d$ can be expressed as

$$F(x_1, \ldots, x_d) = \mathcal{C}(F_1(x_1), \ldots, F_d(x_d)) \quad \text{for any } (x_1, \ldots, x_d) \in \mathbb{R}^d. \tag{12}$$

The function $\mathcal{C}$ is a so-called copula, a distribution function on $[0,1]^d$ with uniform marginals. It is not our intention to discuss copulas in greater detail here; we refer to monographs by Nelsen (1999) or Joe (1997) for further information. McNeil et al. (2005) and Cherubini et al. (2004) contain introductions with a special emphasis to applications in finance. It is sufficient to note that $\mathcal{C}$ is unique if the marginal distributions are continuous. Moreover, combining given marginals with a chosen copula through (12) always yields a multivariate distribution with those marginals. For the purpose of illustration of the methods presented below, we will use copulas of the so-called Frank family. These are defined by

$$\mathcal{C}_\theta(u, v) = -\frac{1}{\theta} \ln\left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}\right), \quad \theta \in [-\infty, \infty],$$

where the cases $\theta = -\infty$, 0 and ∞, respectively, are understood as limits. The choice of the Frank family is merely motivated by its mathematical properties. It is in particular comprehensive, meaning that $\mathcal{C}_\theta$ models a wide class of dependence scenarios for different values of the parameter θ: perfect positive dependence, or comonotonicity (for $\theta = \infty$), positive dependence (for $\theta > 0$), negative dependence (for $\theta < 0$), perfect negative dependence, or countermonotonicity (for $\theta = -\infty$) and independence (for $\theta = 0$).

In the situation of point processes, there are two situations where the copula modelling is particularly useful. First, if the event-time points $T_i(1), \ldots, T_i(d)$ have fixed and continuous distributions, say $F_1, \ldots, F_d$, then choosing some suitable copula $\mathcal{C}_{\boldsymbol{T}}$ yields the distribution F of the d-dimensional event-time point $\boldsymbol{T}_i = (T_i(1), \ldots, T_i(d))$ via (12).

Secondly, the copula approach can be used for constructing multivariate distributions with Poisson marginals (see also Joe (1997) and Nelsen (1987)). Although such distributions may

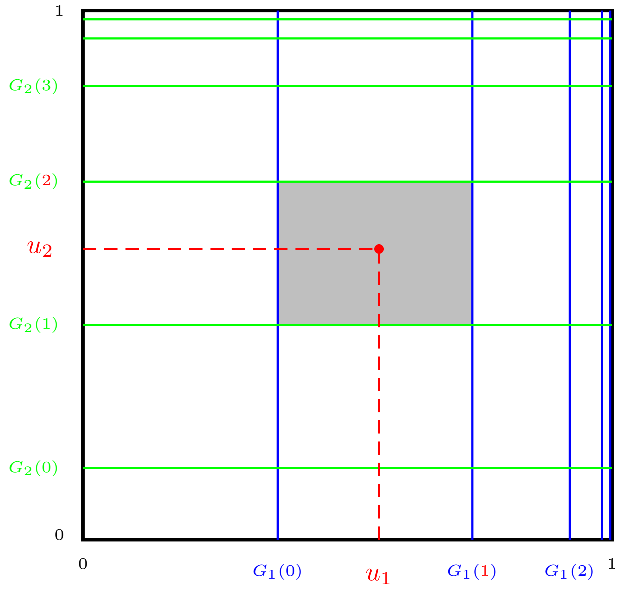

Figure 9: Generation of random variables with Poisson marginals and Frank copula.

not possess nice stochastic interpretations and have to be handled with care because of the non-continuity of the marginals, they cover a wide range of dependence possibilities; see Griffiths et al. (1979), Nešlehová (2004), Pfeifer and Nešlehová (2004) and Denuit and Lambert (2005) for further details. Our focus here lies in describing how the generation of two dependent Poisson random variables using copulas works.

For the moment, suppose G_1 and G_2 denote Poisson distributions and $\mathcal{C}$ a chosen copula. In the first step, we generate a random point (u, v) in the unit square $[0, 1] \times [0, 1]$ from the copula $\mathcal{C}$. Thereafter, we determine integers i and j in a way that (u, v) lies in the rectangle $R_{ij} := (G_1(i-1), G_1(i)] \times (G_2(j-1), G_2(j)]$. Note that the choice of the i and j is unique. The point (i, j) is then the realization of a two dimensional Poisson random vector with copula $\mathcal{C}$ and marginals G_1 and G_2. Figure 9 shows a random generation of a pair (N_1, N_2) with Poisson marginals with parameters 1 and 2 and a Frank copula with parameter -10; the horizontal and vertical lines indicate the subdivision of the unit square into the rectangles R_{ij}. Here for instance, all simulated random points falling into the shaded rectangle generate the (same) pair $(1, 2)$.

4.3 Dependent point processes

In this subsection, we finally present two methods for constructing dependent Poisson point processes. This task however implicitly involves another important question: what does

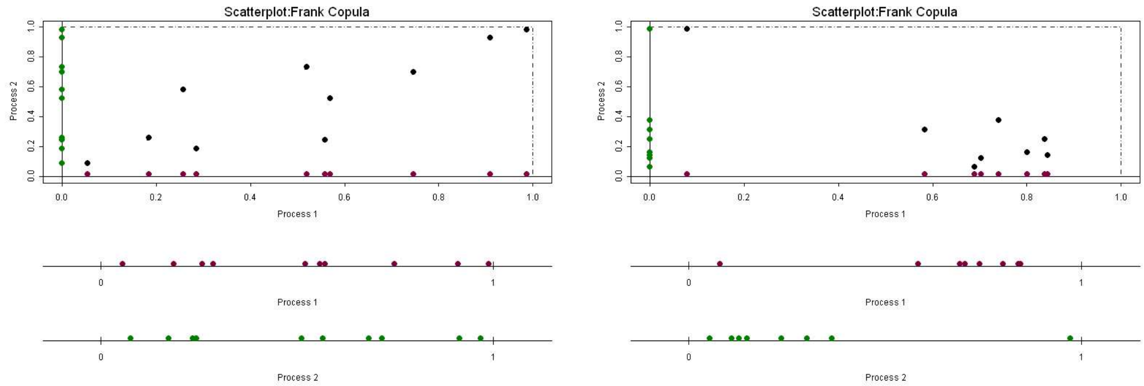

Figure 10: Dependent homogeneous Poisson processes constructed via Method I. The event positions are generated by the Frank copula with parameter 10 (left) and -10 (right).

dependence between point processes mean and how can we describe it? For random variables, there exist several ways of describing dependence. For instance one can calculate dependence measures like linear correlation, rank correlations like Spearman's rho or Kendall's tau, or investigate dependence concepts like quadrant or tail dependence, or indeed one can look for a (the) copula. For stochastic processes, the notion of Lévy copulas offers an interesting alternative if the process is Lévy; see for instance Cont and Tankov (2004), Kallsen and Tankov (2004) and Barndorff-Nielsen and Lindner (2004). Alternative measures of dependence have been proposed for point processes. Griffiths et al. (1979) use the following analogue of the linear correlation coefficient. Suppose ξ_1 and ξ_2 are point processes defined on the same state space, say $[0,\Delta]^d$. Then the correlation between the two processes can be expressed by the correlation coefficient $\rho(\xi_1(A), \xi_2(B))$ between the random variables $\xi_1(A)$ and $\xi_2(B)$ for some sets A and B.

Construction Method I. This method is based upon an extension of (10) and produces Poisson point processes with the same random number N of events. Let $\xi = \sum_{i=1}^{N} I_{\boldsymbol{T}_i}$ be a Poisson process with iid d-dimensional event points $\boldsymbol{T}_i = (T_i(1), \ldots, T_i(d))$ whose joint distribution for each i is given through a copula $\mathcal{C}_{\boldsymbol{T}}$. We can again think of the $T_i(k)$'s being loss occurrence times in d different classes, say. Following (10), the marginal processes $\xi(k) = \sum_{i=1}^{N} I_{T_i(k)}$, $k = 1, \ldots, d$ are Poisson, but dependent.

Figure 10 illustrates Method I. The counting variable N is Poisson with parameter 20 and

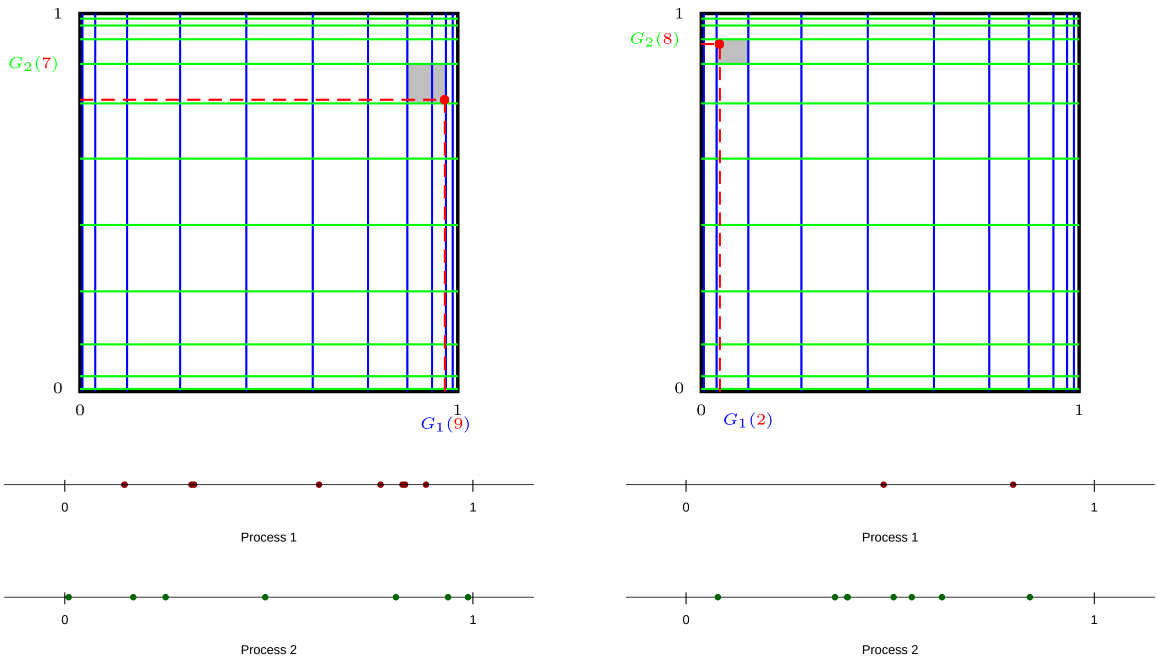

Figure 11: Dependent Poisson point processes constructed with Method II. The total numbers of events are generated by the Frank copula with parameter 10 (left) and -10 (right).

$T_i(k)$, $k = 1, \ldots, d$, are uniform with joint distribution function given by the Frank copula. The resulting dependent Poisson point processes are displayed on the axes as well as under the graphs for a better visualisation. The parameter of the Frank copula is 10 (left) yielding highly positively correlated event time points and -10 (right) producing highly negatively correlated event time points. The loss event times in the left panel for both types cluster in similar time periods, whereas the event times in the right panel tend to "avoid" each other. This is a typical example of what one could call *dependence engineering*.

As shown in Pfeifer and Nešlehová (2004), the correlation of $\xi(k)$ and $\xi(l)$ is given by

$$\rho(\xi(k)(A), \xi(l)(B)) = \frac{F_{kl}(A \times B)}{\sqrt{F_k(A)F_l(B)}}, \quad k, l = 1, \ldots, d, \tag{13}$$

where F_{kl} stands for the joint distribution of $T_i(k)$ and $T_i(l)$ and F_k and F_l denote the marginal distributions of $T_i(k)$ and $T_i(l)$, respectively. Note especially that, since $F_{kl}(A \times B)$ is a probability, the correlation is never negative. Hence, only positively correlated Poisson processes can be generated in this way, the reason being that the marginal processes all have the same number N of events. Construction Method I is thus particularly suitable for situations where the events are triggered by a common underlying effect.

Construction Method II allows for variable numbers of events. Here, we first generate dependent Poisson random variables $N_1, \dots, N_d$ with copula $\mathcal{C}_{\boldsymbol{N}}$, for instance using the modelling approach described in the previous subsection. Secondly, the occurrence time points $T_i(k)$ are again generated as (possibly dependent) margins of a d-dimensional time-event point $\boldsymbol{T}_i = (T_i(1), \dots, T_i(d))$. In this way, we obtain d dependent processes $\xi(k) = \sum_{i=1}^{N_k} I_{T_i(k)}$, $k = 1, \dots, d$. Figure 11 illustrates this method. The occurrence time points are chosen independent and uniformly distributed. The counting variables are Poisson with parameters 5 each and $\mathcal{C}_{\boldsymbol{N}}$ is the Frank copula with parameters 10 (left plot) and -10 (right plot). Hence, the counting variables are strongly positively and negatively dependent, respectively. As a consequence, the choice of the Frank copula with a comparatively strong positive dependence structure ($\theta = 10$) leads to similar number of events for both processes. On the other hand, when ($\theta = -10$), events in both processes tend to avoid each other. This is best seen with Figure 12, which combines Method II with superposition. For each interval $[n-1, n)$ the Poisson processes have been generated independently using Method II with the same parameters as in Figure 11 (right) and joined together to a process on $[0, 6)$. The 6 could correspond to a time horizon of 6 years, say.

In case the $T_i(k)$'s are mutually independent, the correlation in this construction is given by

$$\rho(\xi(k)(A), \xi(l)(B)) = \rho(N_k, N_l)\sqrt{F_k(A)F_l(B)}, \quad k, l = 1, \dots, d; \tag{14}$$

see Pfeifer and Nešlehová (2004). Note that this formula involves the correlation coefficient of the counting variables N_k and N_l. Hence, by a suitable choice of $\mathcal{C}_{\boldsymbol{N}}$ which governs the joint distribution of $N_1, \dots N_d$, a wider range of correlation, in particular negative, is achievable.

Operational loss occurrence processes will typically be more complex than those constructed solely via Methods I or II. In order to come closer to reality, both methods can be combined freely using superposition and/or refined by thinning. For example, Figure 13 shows a superposition of independent homogeneous Poisson point processes with different intensities over $[0, 1]$ with homogeneous but highly positively dependent Poisson point processes generated by Method I as in Figure 10.

A broad palette of models now becomes available, which may contribute to a better understanding of the impact of interdependence between various risk classes on the quantification of the resulting aggregate loss random variables, like the L_k's in Section 3.1. For this we need to include the loss severities explicitly; this step is discussed in the next section.

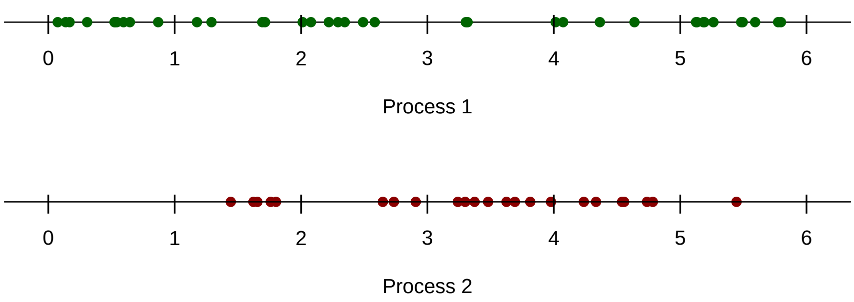

Figure 12: Dependent Poisson point processes on [0,6).

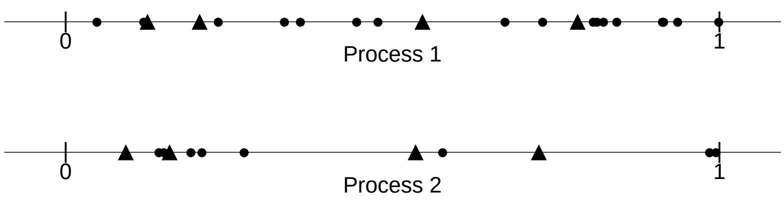

Figure 13: Superposition of independent homogeneous Poisson processes with intensity 10 (Process 1) and 8 (Process 2) over $[0, 1]$ (bullets) and dependent Poisson processes generated by the Frank copula with parameter 20 (triangles).

4.4 Dependent aggregate losses

The loss severities can be included in the point process modelling in a number of ways. For example, we can consider d-dimensional point processes where the first component describes the time and the remaining $d-1$ components the sizes of the reported losses.

For the sake of simplicity, we illustrate some of the modelling issues in the case of stationary and independent loss amounts. Consider two aggregate losses L_1 and L_2, corresponding to two particular operational risk types and some period of time, $[0,\Delta]$ say. As in Subsection 4.1, assume that the loss occurrence times of each risk type form a Poisson point process, $\xi(k)=\sum_{i=1}^{N_k} I_{T_i(k)}$, $k=1,2$, say. The processes $\xi(1)$ and $\xi(2)$ may be dependent and modelled by one of the techniques described in the previous subsection; we discuss several concrete examples below. Furthermore, we denote the severities corresponding to $T_i(1)$ and $T_i(2)$ by $X_i(1)$ and $X_i(2)$, respectively. The severities are each assumed to be iid and $X_i(1)$ and $X_j(2)$ independent of one another for $i\neq j$; hence we only allow for dependence between $X_i(1)$ and $X_i(2)$. Recall that the entire risk processes can be described as point processes according to $\widetilde{\xi}(k)=\sum_{i=1}^{N_k} I_{(T_i(k),X_i(k))}$, $k=1,2$. The corresponding aggregate losses are given by

$$L_1=\sum_{i=1}^{N_1} X_i(1) \quad \text{and} \quad L_2=\sum_{i=1}^{N_2} X_i(2).$$

Note that although the dependence between the loss occurrence processes $\xi(1)$ and $\xi(2)$ very much determines the dependence between L_1 and L_2, the precise location of the loss occurrence times within the time period of interest does not yet enter into the modelling of the aggregate losses explicitly. The results below are hence comparable with those obtained from models which do not directly address the dependence structure between the loss occurrence processes, as for instance in Powojowski et al. (2002) or Frachot et al. (2004).

We now focus on the correlation between L_1 and L_2 for several selected types of dependence between the underlying loss occurrence processes $\xi(1)$ and $\xi(2)$. First, if $\xi(1)$ and $\xi(2)$ are constructed using Method I, we have as in Pfeifer and Nešlehová (2004) that

$$\rho(L_1,L_2)=\frac{\mathrm{E}(X_1(1)X_1(2))}{\sqrt{\mathrm{E}(X_1(1))^2\,\mathrm{E}(X_1(2))^2}}. \tag{15}$$

Note that similarly to (13), the right hand side is never zero nor becomes negative for positive loss amounts. This is different when $\xi(1)$ and $\xi(2)$ are constructed using Method II, for there we have, in case $X_i(1)$ and $X_i(2)$ are independent for any i, that similar to (14),

$$\rho(L_1,L_2)=\rho(N_1,N_2)\frac{\mathrm{E}(X_1(1))\,\mathrm{E}(X_1(2))}{\sqrt{\mathrm{E}(X_1(1))^2\,\mathrm{E}(X_1(2))^2}}; \tag{16}$$

see again Pfeifer and Nešlehová (2004). As the correlation is driven by the correlation of the counting variables N_1 and N_2, it can be negative if the losses corresponding to different risk types are caused by effects which rather do not occur simultaneously. Note also that (16) coincides with the result obtained by Frachot et al. (2004).

Finally, we would like to mention one particularly simple special case of superposition. Assume that the time occurrence processes are generated as sums of independent homogeneous Poisson point processes ξ_k with intensities λ_k, $k = 1, 2, 3$ in the sense that $\xi(1) = \xi_1 + \xi_3$ and $\xi(2) = \xi_2 + \xi_3$. Then (16) leads to

$$\rho(L_1, L_2) = \left(\frac{\lambda_3}{\sqrt{(\lambda_1 + \lambda_3)(\lambda_2 + \lambda_3)}} \right) \frac{\mathrm{E}(X_1(1))\,\mathrm{E}(X_1(2))}{\sqrt{\mathrm{E}(X_1(1))^2\,\mathrm{E}(X_1(2))^2}}. \tag{17}$$

This model corresponds to the setup considered by Powojowski et al. (2002) and allows for variable *positive* correlation.

As noted in Frachot et al. (2004), the correlation coefficient of the aggregated losses does depend on the loss severity distribution in a way which yields comparatively small values of ρ for heavy-tailed marginal loss distributions. This fact is however due rather to the properties of the correlation coefficient itself and does not necessarily imply lack of dependence; see Embrechts et al. (2002) and especially McNeil et al. (2005), Example 5.26. Consequently, even if the value of $\rho(L_1, L_2)$ is close to zero, VaR_α of the sum $(L_1 + L_2)$ (or more general of $\sum_{k=1}^{d} L_k$) can be substantially different from VaR_α in the case of independent (aggregated) loss random variables L_k. A more accurate study on the impact of the dependence between the risk processes on the risk capital however calls for further detailed research and larger data sets.

Apart from the simplified situation of stationary and independent loss amounts, the methods presented in Section 4.3 can moreover be used in much more complex dependence modelling. In particular, dependent loss processes can readily be constructed as to allow for non-stationarity in the loss severity distribution and/or non-stationarity in the dependence structure between loss severities. Modelling dependence between two or more loss processes however still remains a delicate and complex issue and definitely warrants more research (especially on the statistical side) before practical guidelines for specific applications can be given. There is a flurry of mathematical research ongoing on this topic; beyond the references already given, see also the *common shock model* by Lindskog and McNeil (2003) or Bäuerle and Grübel (2005). The latter paper also discusses the construction of dependent loss processes in a point process context.

5 Conclusion

From a mathematical point of view, the capital charge calculation for Operational Risk within the Basel II AMA corresponds to the calculation of risk capital of the form

$$\rho_\alpha\left(\sum_{k=1}^{d} L_k\right) = \rho_\alpha(L). \tag{18}$$

for some risk measure ρ_α at confidence level α. For $k = 1, \ldots, d$,

$$L_k = \sum_{i=1}^{N_k} X_i(k) \tag{19}$$

denotes the aggregate loss for loss type k with loss frequency (N_k) and severity $(X_i(k))$ random variables. Under Basel II $\rho = \mathrm{VaR}$ and $\alpha = 0.999$. The latter implies that modelling of $\mathsf{P}(L > x)$ for x large is needed. Due to the lack of information on the joint distribution $(L_1, \ldots, L_d)$, various shortcuts for the calculation of (18) are in use. One widely uses a two-step procedure: first calculate $\sum_{k=1}^{d} \mathrm{VaR}_{99.9\%}(L_k)$, which only uses the marginal dfs of $L_1, \ldots, L_d$, and then reduce this capital charge measure by some correlation considerations based on some dependence assumptions in (19). This approach raises several issues.

- The fact that α is large calls for some form of Extreme Value Theory to be used. At the moment, and based on the recently available data, no clean standardized EVT approach is available. We provide a generalization of the POT method allowing for non-stationarity in the frequency and severity of the losses.

- The fact that a joint model for $(L_1, \ldots, L_d)$ is not known, we give an optimization example that allows to calculate by how much $\sum_{k=1}^{d} \mathrm{VaR}_\alpha(L_k)$ may even underestimate the true risk measure $\mathrm{VaR}_\alpha(L)$. Whether this situation actually occurs in practice is an important issue for further investigation.

- We finally show how the theory of point processes offers a natural environment for the construction of dependence scenarios on $(L_1, \ldots, L_d)$ and in particular derive some results used in the operational risk literature as special cases.

The above methodological tools are tested on some examples, including a $d = 3$ data set of operational risk losses. This example is given mainly to highlight the practicality of the techniques introduced. As the paper stands, we want to stress relevant areas for future research on topics which we believe will have an important impact on quantitative risk measurement in general and on the quantitative (AMA) modelling of operational risk more in particular.

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. They also thank Giovanni Puccetti for the example reported in Section 3. The numerous comments by two anonymous referees led to a substantial improvement of the paper.

References

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

Balkema, G. and Embrechts, P. (2006). *High Risk Scenarios and Extremes. A Geometric Approach.* Book manuscript, Zurich Lectures in Advanced Mathematics, European Mathematical Society, to appear.

Barndorff-Nielsen, O. and Lindner, A. (2004). Some aspects of lévy copulas. Preprint, Technische Universität München.

Bäuerle, N. and Grübel, R. (2005). Multivariate counting processes: copulas and beyond. preprint, University of Hannover.

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

Bielecki, T. and Rutkowski, M. (2002). *Credit Risk: Modeling, Valuation, and Hedging.* Springer, Berlin.

Brown, D. and Wang, J. (2005). Discussion on "Quantitative Models for Operational Risk: Extremes, Dependence and Aggregation". Presentation. *Implementing an AMA to Operational Risk*, Federal Reserve Bank of Boston, May 18-20, http://www.bos.frb.org/bankinfo/conevent/oprisk2005/.

Chavez-Demoulin, V. (1999). *Two Problems in Environmental Statistics: Capture-Recapture Analysis and Smooth Extremal models.* PhD thesis, Department of Mathematics, Swiss Federal Institute of Technology, Lausanne.

Chavez-Demoulin, V. and Davison, A. (2005). Generalized additive models for sample extremes. *Journal of the Royal Statistical Society, Series C*, 54(1):207–222.

Chavez-Demoulin, V. and Embrechts, P. (2004). Smooth extremal models in finance. *The Journal of Risk and Insurance*, 71(2):183–199.

Cherubini, U., Luciano, E., and Vecchiato, W. (2004). *Copula Methods in Finance.* Wiley, Chichester.

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

Cont, R. and Tankov, P. (2004). *Financial Modelling with Jump Processes.* Chapman & Hall/CRC, London.

Cruz, M. (2002). *Modelling, Measuring and Hedging Operational Risk.* Wiley, Chichester.

Cruz, M., editor (2004). *Operational Risk Modelling and Analysis: Theory and Practice.* Risk Waters Group, London.

Daley, D. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods.* Springer, New York, 2. edition.

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/.

Denuit, M. and Lambert, P. (2005). Constraints on concordance measures in bivariate discrete data. *Journal of Multivariate Analysis*, 93:40–57.

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., editors, *Advances in Computational Finance*, pages 3–12. Kluwer Academic Publishers, Amsterdam. Reprinted in Journal of Risk Finance, 1 (Winter 2000), 30-36.

EBK (2005). *Basel II, Umsetzung in der Schweiz. Entwurf September 2005.* Swiss Federal Banking Commission.

Embrechts, P., editor (2000). *Extremes and Integrated Risk Management.* Risk Waters Group, London.

Embrechts, P., Furrer, H., and Kaufmann, R. (2003). Quantifying regulatory capital for operational risk. *Derivatives Use, Trading & Regulation*, 9(3):217–233.

Embrechts, P., Hoeing, A., and Puccetti, G. (2005). Worst VaR secenarios. *Insurance: Mathematics and Economics*, to appear.

Embrechts, P., Kaufmann, R., and Samorodnitsky, G. (2004). Ruin theory revisited: stochastic models for operational risk. In Bernadell, C., Cardon, P., Coche, J., Diebold, F., and Manganelli, S., editors, *Risk Management for Central Bank Foreign Reserves*, pages 243–261. ECB, Frankfurt.

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 Univ. Press, Cambridge.

Embrechts, P. and Puccetti, G. (2006a). Aggregating risk capital. with an application to operational risk. *The Geneva Risk and Insurance Review*, to appear.

Embrechts, P. and Puccetti, G. (2006b). Bounds for functions of dependent risks. *Finance and Stochastics*, to appear.

Falk, M., Hüsler, J., and Reiss, R. (2004). *Laws of Small Numbers: Extremes and Rare Events.* Birkhäuser, Basel, 2. edition.

Frachot, A., Roncalli, T., and Salomon, E. (2004). The correlation problem in operational risk. Preprint, Crédit Lyonnais, Paris.

Green, P. and Silverman, B. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach.* Chapman and Hall, London.

Griffiths, R., Milne, R. K., and Wood, R. (1979). Aspects of correlation in bivariate Poisson distributions and processes. *Australian Journal of Statistics*, 21(3):238–255.

Joe, H. (1997). *Multivariate Models and Dependence Concepts.* Chapman & Hall, London.

Kallsen, J. and Tankov, P. (2004). Lévy copulas for general lévy processes. Preprint, Technische Universität München.

King, J. (2001). *Measurement and Modelling Operational Risk.* Wiley.

Kingman, J. (1993). *Poisson Processes.* Clarendon Press, Oxford.

Leadbetter, M. (1991). On a basis for 'peaks over threshold' modeling. *Statistics and Probability Letters*, 12(4):357 – 362.

Lindskog, F. and McNeil, A. (2003). Common Poisson shock models: Application to insurance and credit risk modelling. *ASTIN Bulletin*, 33:209–238.

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

McNeil, A., 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.

Nelsen, R. B. (1987). Discrete bivariate distributions with given marginals and correlation. *Communications in Statistics - Simulation*, 16(1):199 – 208.

Nelsen, R. B. (1999). *An Introduction to Copulas.* Springer, New York.

Nešlehová, J. (2004). *Dependence of Non-Continuous Random Variables.* PhD thesis, Carl von Ossietzky Universität Oldenburg.

Pfeifer, D. and Nešlehová, J. (2004). Modeling and generating dependent risk processes for IRM and DFA. *ASTIN Bulletin*, 34(2):333–360.

Powojowski, M. R., Reynolds, D., and Tuenter, H. J. H. (2002). Dependent events and operational risk. *Algo Research Quarterly*, 5(2):68–73.

Puccetti, G. (2005). *Dependency Bounds for Functions of Univariate and Multivariate Risks.* PhD thesis, University of Pisa.

Rachev, S., Menn, C., and Fabrizzi, F. (2005). *Fat-Tailed and Skewed Asset Return Distributions.* Wiley, New York.

Rachev, S. and Rüschendorf, L. (1998a). *Mass Transportation Problems. Volume I: Theory.* Springer, N.Y.

Rachev, S. and Rüschendorf, L. (1998b). *Mass Transportation Problems. Volume II: Applications.* Springer, New York.

Reiss, R.-D. (1993). *A Course on Point Processes.* Springer, New York.

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

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