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/`

Draft Version, June 15, 2005

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

[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

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. A fast growing literature exists on the various aspects of operational risk modelling; see the list of references towards the end of the paper. For a textbook discussion very much in line with our paper, see 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, as is well known, not all operational risk data lend themselves easily to a full quantitative analysis. The analytic methods discussed cover a broad range of issues which will typically enter in the development of an advanced measurement approach, AMA in the language of Basel II.

In Section 2, we first present some more advanced techniques from the realm of extreme value theory (EVT). EVT is considered as a canonical set of tools for analyzing rare events; several of the operational risk classes exhibit properties which very naturally call for an EVT analysis. Especially the non-stationarity of most long-term operational risk data however warrants an approach "beyond classical EVT".

In Section 3, we turn to the problem of modelling the interdependencies 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. Whereas the results from Section 2 are immediately applicable (as will be shown on some data), the techniques of Section 3 are presented in order to offer a first glimpse on what may be obtained. We expect that more results of this type will become available in the near future.

In Section 4 we leave the detailed modelling of loss processes and turn to the question of how to combine or aggregate risk measures across several operational risk classes when no precise dependence information is available. This leads to well-known optimization problems known under the names Fréchet class problems or mass transportation problems. Also here, the notion of copula comes in useful. The techniques discussed in this section can also be used to tackle the problem of risk aggregation between risk classes of different types, as there are market, credit, operational and underwriting risk for instance.

A final Section 5 contains some conclusions and thoughts on further research.

2 Advanced EVT Models

2.1 Why EVT?

The key attraction of EVT is that it offers a set of ready-made approaches to the most challenging problem of quantitative 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. (2003a) 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. (2003a) 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 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. However, even in these cases, one may have to go beyond standard EVT to come up with a correct modelling. To illustrate the latter issue, consider Figure 1 taken from Embrechts et al. (2004). For our purposes, it suffices to recall that the data span a 10 year period for three different operational risk loss types, referred to as Types 1, 2 and 3. The stylised facts observed here are:

- the historical period is relatively short (only 10 years of data);
- loss amounts very clearly show extremes;
- loss occurrence times are irregularly spaced in time, and
- the number of occurrences (though relatively few) seems to increase over time with a radical change around 1998.

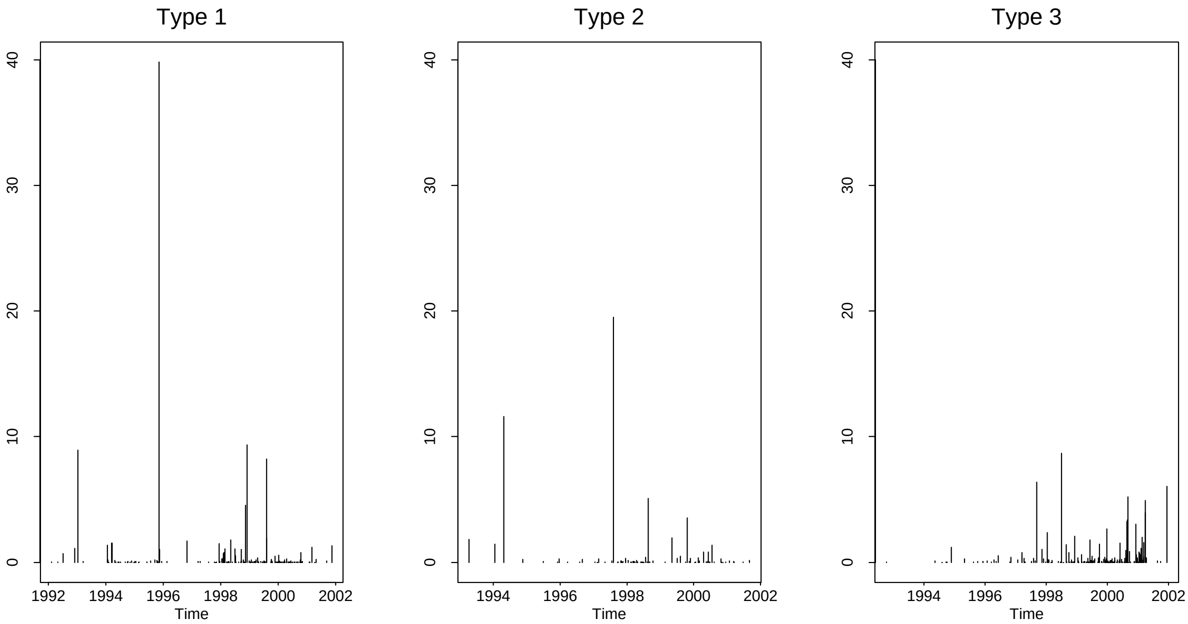

Figure 1: Operational risk losses. From left to right: Type 1 ($n = 162$), Type 2 ($n = 80$), Type 3 ($n = 175$).

The last point very clearly highlights the presence of non-stationarity in current operational loss data. The "discontinuity" may be due to the effort to build such a database of losses of the same type going back about 10 years; quantifying 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 effects, management action, M&A) or changes in the economic/political/regulatory environment in which the company operates (exogenous effects).

In this section, we adapt classical EVT to take both non-stationarity and covariate modelling (different types of losses) into account. This section should be viewed as a first illustrative example of these techniques. Chavez-Demoulin (1999), Chavez-Demoulin and Davison (2005) contain the relevant methodology. The latter paper explains the new technique with finance and insurance related applications in mind. 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. Subsection 2.4 discusses some issues resulting from the modelling of very heavy-tailed data.

2.2 The basic EVT methodology

Over the recent years, EVT has been recognized as a very 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 insurance and finance applications in mind. 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.

Below, we only 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, 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). 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. For operational loss modelling one typically finds $\kappa > 0$ which corresponds to ground-up losses $Z_1, \ldots, Z_q$ 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) reported 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 $Z_1, \ldots, Z_q$ 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 a further step, the POT method can be extended by allowing the parameters λ, σ, κ to be dependent on time and explanatory variables allowing for non-stationarity; this is useful for applications to operational risk modelling. 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 obvious "increase" in loss intensity in Figure 1. We moreover will differentiate between the different loss types and adjust the parameters κ and σ accordingly.

Before we proceed with the data analysis, we briefly review the main risk measures to be analysed throughout this paper, Value-at-Risk (VaR) and Expected-Shortfall (ES) (also referred to as "conditional VaR", "mean excess loss", "beyond VaR" or "tail VaR"). The ES is an alternative risk measure that has been proposed to alleviate some conceptual problems inherent in VaR. For α close to 1 (0.999, say) and a general loss random variable X with continuous distribution function F, these measures are defined as follows:

$$\mathrm{VaR}_\alpha = F^{-1}(1 - \alpha),$$

$$\mathrm{ES}_\alpha = E(X \mid X > \mathrm{VaR}_\alpha).$$

We refrain from discussing the various issues underlying the choice and definition of these risk measures, as there are: the precise definition of F^{-1} or the problem in the definition of ES_α whenever X is non-continuous. The latter may lead to the loss of sub-additivity of ES_α, a

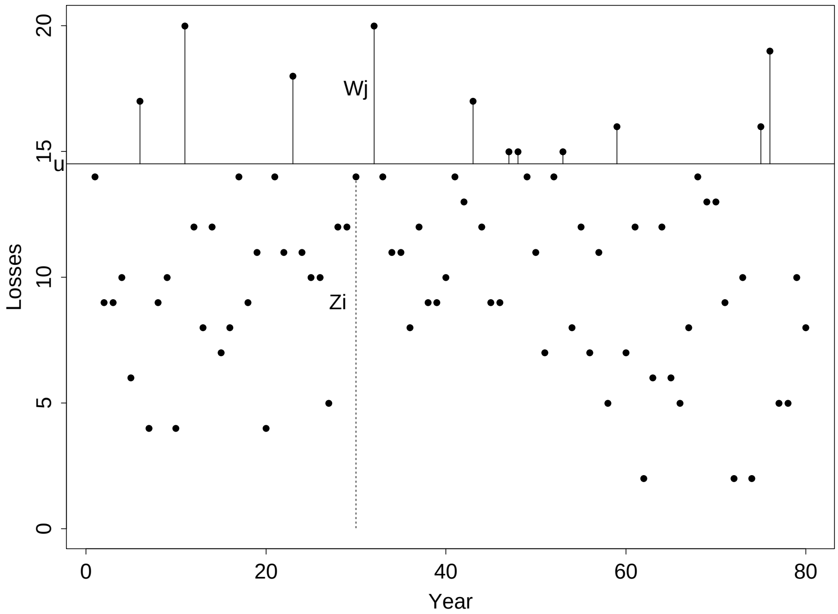

Figure 2: The point process of exceedances (POT).

problem which easily can be remedied; see McNeil et al. (2005) for a discussion and further references. We assume throughout that VaR_α and ES_α as given above are well defined. In particular for ES_α this means that $\mathrm{E}(X) < \infty$, so that in the GPD case, $\kappa < 1$.

In cases where the POT method can be applied, for given u, these measures can be estimated as follows:

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

and

$$\widehat{\mathrm{ES}}_\alpha = \left\{ \frac{1}{1-\hat{\kappa}} + \frac{\hat{\sigma} - \hat{\kappa} u}{(1-\hat{\kappa})\widehat{\mathrm{VaR}}_\alpha} \right\} \widehat{\mathrm{VaR}}_\alpha. \tag{4}$$

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 into the freeware EVIS by Alexander McNeil, available under www.math.ethz.ch/~mcneil.

Though an analysis of the data in Figure 1 is self-contained, the interested reader wanting to learn more about the specifics of modelling non-stationarity and covariates into the POT method is adviced to read Chavez-Demoulin and Embrechts (2004) and the references therein before proceeding. The less technical reader will no doubt find the analysis presented in the next section sufficiently easy to follow in order to grasp the relevance of this more advanced EVT method.

2.3 POT analysis of the operational loss data

In the previous subsections, we briefly laid the foundation of the approach towards the analysis of extremes based on the exceedances of a high threshold. We now return to the operational risk data of Figure 1 which consists of three different types over a 10 year period. Our analysis below is more illustrative; in order to become fully applicable, much larger operational loss data bases will have to become available. From the discussion of the data, it follows that we should at least take the risk type τ as well as the non-stationarity (switch around 1998) into account. First, pool the data in the three panels of Figure 1. 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 regime switching, 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. A result from that analysis is that for these data, small variations in the value of the threshold have nearly no impact. Given sufficient data, much more than in Figure 1, our method would allow to model VaR_α and ES_α as functions of time: are they constant or changing in time? Are they dependent on the type of losses?

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 or regime switch and

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

Of course a more formal test on the existence and value of t_c can be included; the rather pragmatic choice of $t_c = 1998$ suffices for this first illustrative analysis. 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.

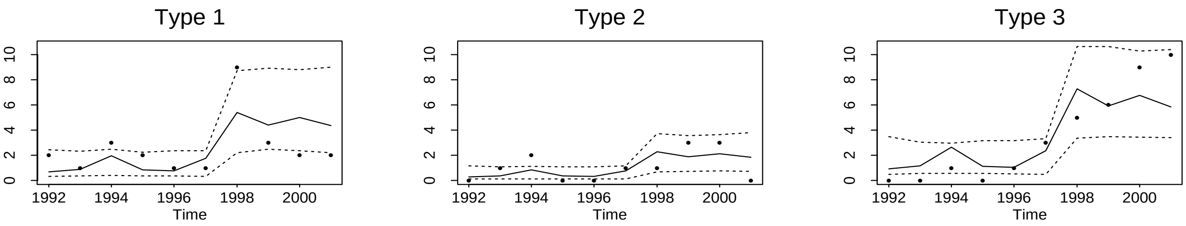

Figure 3: Operational risk losses. From left to right: Estimated Poisson intensity $\hat{\lambda}$ and 95% confidence intervals for data of loss type 1, 2, 3. The points are the yearly numbers of exceedances over $u = 0.4$.

The selected model for the Poisson intensity $\lambda(t, \tau)$ is

$$\log \hat{\lambda}(t, \tau) = \hat{\gamma}_\tau I_\tau + \hat{\beta} I_{(t>t_c)} + \hat{g}(t). \tag{5}$$

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 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)). 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 global increase of the estimated intensity curves therefore seems to be in accordance with reality. Note that the inclusion of the time dependent function $g(t)$ allows us to model this non-stationarity. The advantage of such a non-parametric technique becomes very clear. It 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 loss-size 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. The shape parameter κ (upper panels) is around

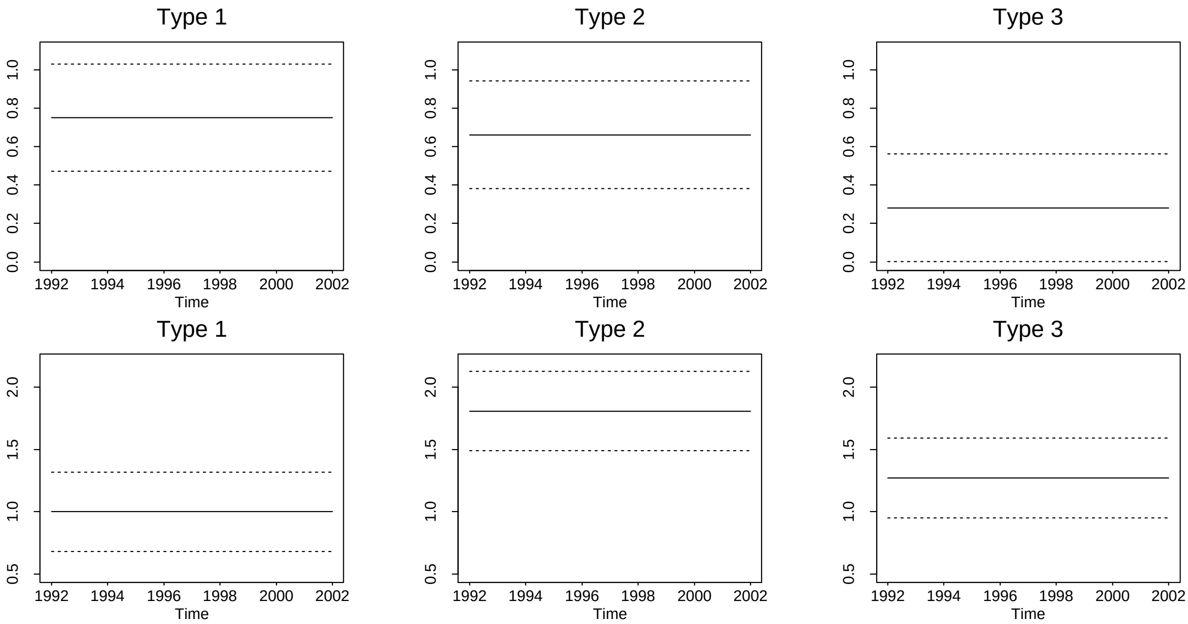

Figure 4: Operational risk losses. Estimated GPD parameters: upper $\hat{\kappa}$, lower $\hat{\sigma}$ and 95% confidence intervals for different loss types.

0.7 for types 1 and 2 (finite mean and infinite variance) and is significantly smaller for type 3 with an estimated value of around 0.3 (finite third moment); this suggests a loss distribution for type 3 with 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 more difficult to estimate than λ, the lack of sufficient data makes the detection of any trend and/or periodic components difficult.

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

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

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 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

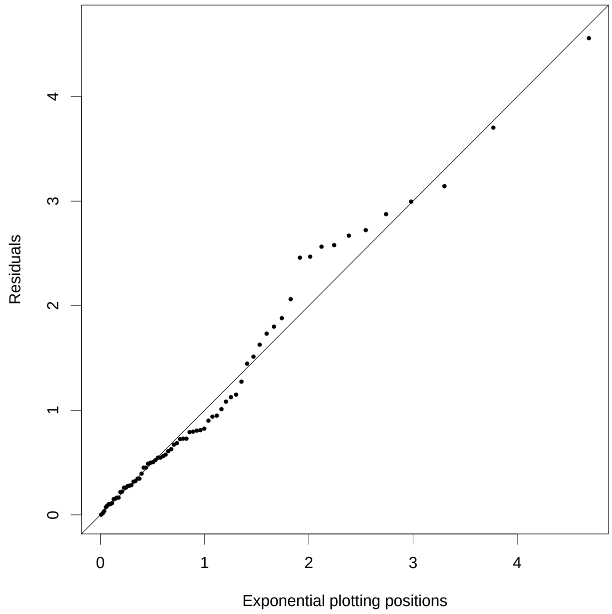

Figure 5: Operational risk losses. Residuals against exponential plotting positions.

the data, a feature which is becoming more and more crucial, particularly in the context of operational and credit risk. Using the estimated parameters $\hat{\lambda}, \hat{\kappa}, \hat{\sigma}$ it is possible to estimate VaR, ES or other the risk 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 estimate procedure at the 99.9 % confidence level, hence we decided not to include such an analysis.

2.4 The one loss causes ruin problem

We further want to make some comments about loss portfolios where the (iid, say) losses follow a Pareto-type distribution with index $1/\kappa$. Based on the concept of Lorenz curve in economics, in Embrechts et al. (1997, Section 8.2), a large claim index is introduced explaining which percentage of the individual losses constitutes a certain percentage of the total portfolio loss. For instance, the famous $20 - 80$ rule corresponds to $1/\kappa = 1.4$. I.e., in an iid Pareto portfolio with index 1.4, 20 % of the individual losses produce 80 % of the total portfolio loss. For $1/\kappa = 1.01$ (a model with still finite mean, but only just) we have a $0.1 - 95$ rule, i.e. 0.1 % of the losses is responsible for 95 % of the total loss amount. In such models (and definitely for $\kappa > 1$) 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 is also to be found in Mandelbrot and Hudson (2004). We will come back to this issue in Subsection 4.3.

3 Dependent Risk Processes

3.1 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. According to the Basel Committee, operational events are classified into distinct business lines (8) and risk types (7). This leads to maximally 56 classes, though some larger banks collect data even for a higher number of cells, sometimes over 100. For each of these cells, 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 following Neslehova and Pfeifer (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 the previous section 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). 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.

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 more information on the topic of point processes, we refer to Chapter 5 in Embrechts et al. (1997), Reiss (1993), Kingman (1993) or the comprehensive monograph by Daley and Vere-Jones (1988).

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 (see Subsection 3.4 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 construct the easiest *point process*:

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

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

$$\xi_n(A) := \sum_{i=1}^{n} I_{T_i}(A) \tag{6}$$

counts the number of losses in the observation period $A \subset [0,\Delta]$. There are several ways in which we can generalize (6) 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 random measure

$$\xi_N := \sum_{i=1}^{N} I_{\boldsymbol{T}_i}. \tag{7}$$

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

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

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 (7), or replaced by $(\boldsymbol{T}_i, \boldsymbol{X}_i)$ where the $\boldsymbol{X}_i$'s denote the corresponding severities (see Subsection 3.4). 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 3.3. 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 (7) 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 (7). 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, respectively. Each of the projections or, marginal processes,

$$\xi(k) = \sum_{i=1}^{N} I_{T_i(k)}, \quad k = 1, \ldots, d, \tag{8}$$

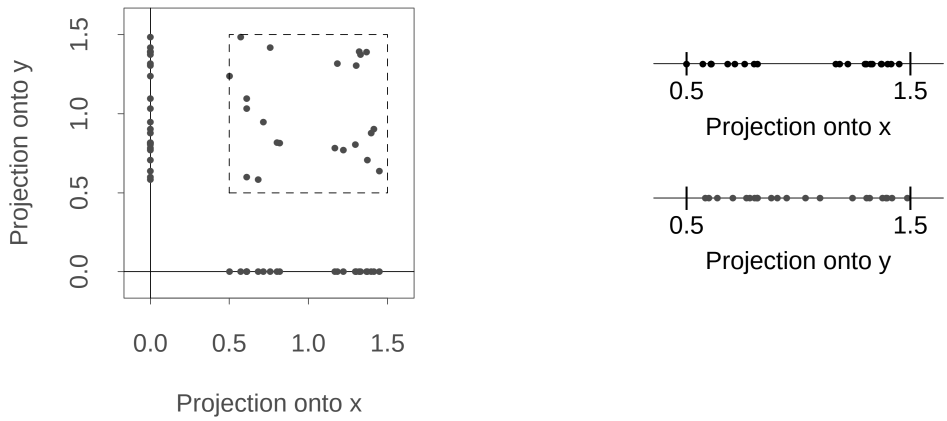

Figure 6: Projections of a two dimensional homogeneous Poisson point process on $[0.5, 1.5] \times [0.5, 1.5]$.

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 6 (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 6 (right).
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 (7), 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 $\mathrm{E}\,\xi = \mathrm{E}\,\xi_2 + \mathrm{E}\,\xi_1$.

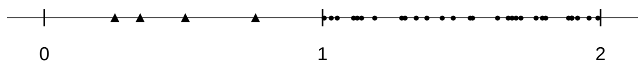

Figure 7: Superposition of a homogeneous Poisson process with intensity 5 over $[0, 1]$ and a homogeneous Poisson process with intensity 20 over $[1, 2]$.

Figure 7 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 11). 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 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

$$\xi_1 := \sum_{i=1}^{N} \varepsilon_i \cdot I_{\boldsymbol{T}_i} \quad \text{and} \quad \xi_2 := \sum_{i=1}^{N} (1 - \varepsilon_i) \cdot I_{\boldsymbol{T}_i}. \tag{9}$$

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.

3.2 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

$$F(x_1, \ldots, x_d) = \mathcal{C}(F_1(x_1), \ldots, F_d(x_d)) \quad \text{for any } (x_1, \ldots, x_d) \in \mathbb{R}^d. \tag{10}$$

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 and insurance. 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 (10) 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 (for $\theta = \infty$), positive dependence (for $\theta > 0$), negative dependence (for $\theta < 0$), perfect negative dependence (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

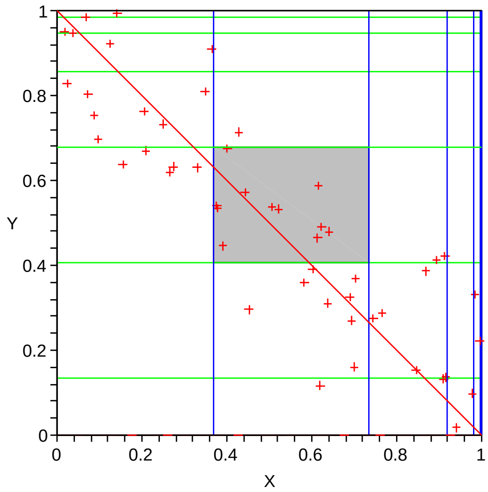

Figure 8: Generation of random variables with Poisson marginals and Frank copula.

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 (10).

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 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), Neslehova (2004), Neslehova and Pfeifer (2004) and Denuit and Lambert (2005) for further details. Our focus here lies in describing how the generation of two 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 8 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 $(1, 2)$.

3.3 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 dependence between point processes mean and how can we describe it? For random variables, there exist several ways of describing dependence precisely. 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 processes, however, a mathematical formulation and well developed theory of dependence and resulting measures do not really exist. There do however exist some partial attempts. Griffiths et al. (1979) use the following analogue of the linear correlation coefficient. Suppose ξ_1 and ξ_2 are one-dimensional 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 (8) 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 $T_i(k)$ being loss occurrence times in d different classes, say. Following (8), the marginal processes $\xi(k) = \sum_{i=1}^{N} I_{T_i(k)}$, $k = 1, \ldots, d$ are Poisson, but dependent. Figure 9 illustrates Method I. The counting variable N is Poisson with parameter 20 and $T_i(k)$, $k = 1, \ldots, d$, 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 we call *dependence engineering.*

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

$$\rho(\xi(k)(A), \xi(l)(B)) = \frac{F_{kl}(A \times B)}{\sqrt{F_k(A) F_l(B)}}, \quad k, l = 1, \ldots, d, \tag{11}$$

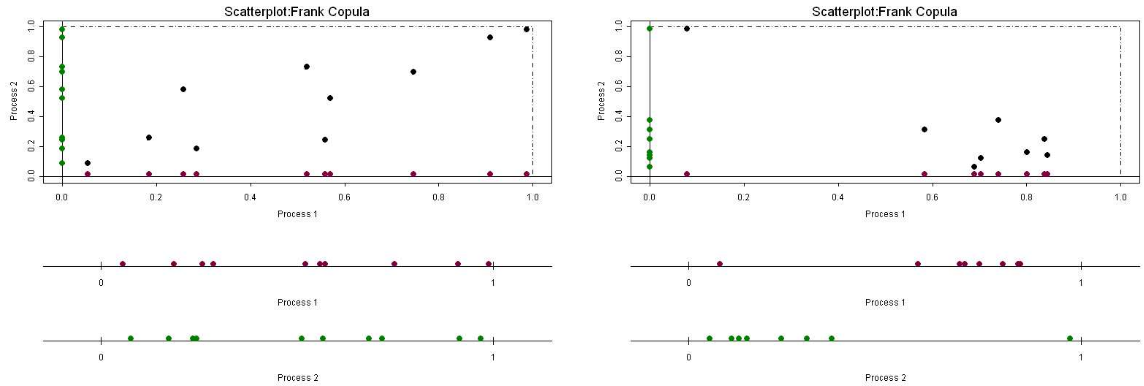

Figure 9: Dependent homogeneous Poisson processes with events generated by the Frank copula with parameter 10 (left) and -10 (right).

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 $N_1, \ldots, N_d$ 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 $T_i = (T_i(1), \ldots, T_i(d))$. In this way, we obtain d dependent processes $\xi(k) = \sum_{i=1}^{N_k} I_{T_i(k)}$, $k = 1, \ldots, d$. Figure 10 illustrates this method. The occurrence time points are chosen independent and uniformly distributed. The counting variables are Poisson with parameters 1 and 2, respectively and $\mathcal{C}_{\boldsymbol{N}}$ is the Frank copula with parameter -10. Hence, the counting variables are strongly negatively correlated. Figure 10 also combines Method II with superposition. For each interval $[n-1, n)$ the Poisson processes have been generated independently and joined together to a process on $[0, 13)$. The 13 could correspond to a time horizon of 13 years, say. Note that, by the choice of the Frank copula with a comparatively strong negative dependence structure ($\theta = -10$), events in both processes tend to avoid each other.

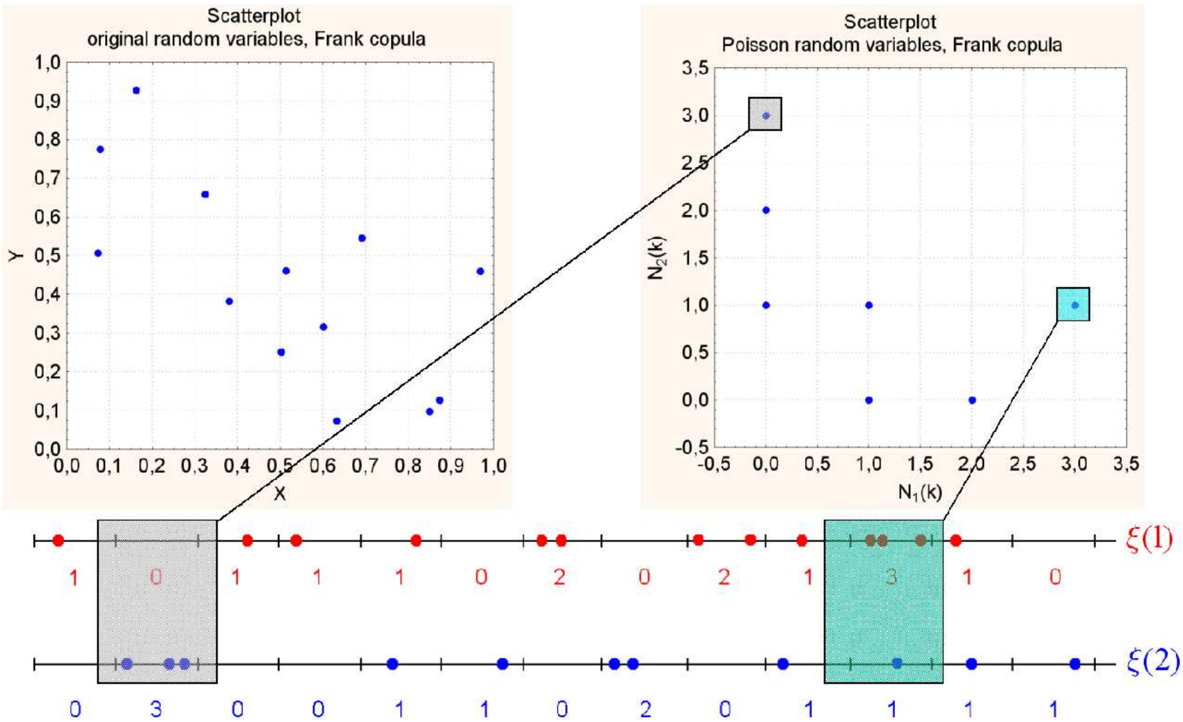

Figure 10: Dependent Poisson point processes on $[0, 13)$.

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

$$\rho(\xi(k)(A), \xi(l)(B)) = \rho(N_k, N_l)\sqrt{F_k(A)F_l(B)}, \quad k, l = 1, \ldots, d; \tag{12}$$

see Neslehova and Pfeifer (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 $N_1, \ldots N_d$, 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 11 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 9.

A broad palette of models now becomes available, which may contribute to a better understanding of the impact of interdependencies between various risk classes on the quantification of operational risk. Needless to say that these dependence engineering constructions are applicable to other types of financial and insurance risk, in particular to credit risk.

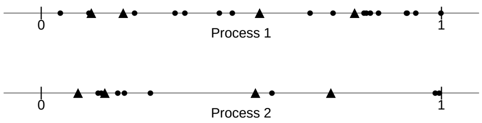

Figure 11: Superposition of independent homogeneous Poisson processes with intensity 10 (Process 1) and 8 (Process 2) over $[0, 1]$ (bullets) and dependent Poisson processes generated by the Frank copula with parameter 20 (triangles).

We conclude this section with the application of the methods discussed above to aggregate losses.

3.4 Dependent aggregate losses

The loss amounts 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. This approach may be particularly useful when there is evidence for non-stationary loss severities. For further details we again refer to the literature given at the beginning of this section.

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 3.1, 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 $i \neq j$. 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 Neslehova and Pfeifer (2004), that

$$\rho(L_1, L_2) = \frac{\mathrm{E}(X_1(1)X_1(2))}{\sqrt{\mathrm{E}(X_1(1))^2\,\mathrm{E}(X_1(2))^2}}. \tag{13}$$

Note that similarly to (11), 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 (12),

$$\rho(L_1, L_2) = \rho(N_1, N_2)\frac{\mathrm{E}(X_1(1))\,\mathrm{E}(X_1(2))}{\sqrt{\mathrm{E}(X_1(1))^2\,\mathrm{E}(X_1(2))^2}}; \tag{14}$$

see Neslehova and Pfeifer (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 mutually exclusive effects. Note also that (14) 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 (14) leads to

$$\rho(L_1, L_2) = \left(\frac{\lambda_3}{\sqrt{(\lambda_1 + \lambda_3)(\lambda_2 + \lambda_3)}}\right)\frac{\mathrm{E}(X_1(1))\,\mathrm{E}(X_1(2))}{\sqrt{\mathrm{E}(X_1(1))^2\,\mathrm{E}(X_1(2))^2}}. \tag{15}$$

This model corresponds to the setup considered by Powojowski et al. (2002) and allows for variable *positive* correlation. The above discussion shows that modelling dependence between two or more loss processes is a delicate and complex issue and definitely warrants more research 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) which also discusses the construction of dependent loss processes in a point process context.
In the absence of precise dependence information, we are forced to come up with a the combination of marginal risk measures into a global risk assessment; the so-called *bottom-up approach*. Methods for handling this type of aggregation problem are discussed in the next section.

4 Aggregating (Operational) Risk Measures

4.1 The risk aggregation problem

A skeletal version of 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 $L_1, \dots, L_d$ 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 compound-sum type as discussed in Section 3. The total loss therefore to be modelled is $L = \sum_{i=1}^{d} L_i$; this random variable in general is very complex as it may contain components with rather different frequency as well as severity characteristics. Moreover, the interdependence between the various L_i's is largely unknown, leading to the dependence engineering issues discussed in the previous section.

Next, a risk measure ϱ is given, mapping L into $\varrho(L)$, the regulatory capital underlying the risky position L. In our paper so far we used $\varrho = \mathrm{VaR}_\alpha$ or $\varrho = \mathrm{ES}_\alpha$ with $\alpha = 0.999$, say. The issue now becomes that one may know $\varrho(L_i)$, $i = 1, \dots, d$, but needs to estimate $\varrho(L)$ solely on the basis of this information. One way forward and indeed supported by the Basel II guidelines, is to put

$$\varrho(L) := \sum_{i=1}^{d} \varrho(L_i). \tag{16}$$

In doing so, one typically assumes that the risks L_i are *comonotone*, i.e. there exists a random variable Y and increasing functions $f_1, \dots, f_d$ so that $L_i = f_i(Y)$, $i = 1, \dots, d$. Indeed, for $L_1, \dots, L_d$ comonotonic and $\varrho = \mathrm{VaR}_\alpha$, (16) holds; see Embrechts et al. (2003b). If ϱ is a coherent risk measure in the language of Artzner et al. (1999), then always

$$\varrho\Big(\sum_{i=1}^{d} L_i\Big) \le \sum_{i=1}^{d} \varrho(L_i), \tag{17}$$

so that $\varrho(L)$ defined in (16) always yields an upper bound of the risk capital required.

As discussed in McNeil et al. (2005), there are essentially three ways in which coherence (or more particularly sub-additivity as in (17)) can break down:

1. Even when the marginal distribution functions $F_{L_i}(x) = \mathsf{P}(L_i \le x)$ are smooth, log-normal, say, one can always construct a dependence structure (i.e. a copula) so that sub-additivity (i.e. (17)) fails.

2. Also in the case of independent L_i's, sub-additivity may fail for very heavy-tailed (i.e. Pareto type) risks.

3. Sub-additivity may further fail because of heavy skewness of the F_{L_i}'s.

For operational risk data, especially 2. and 3. above are particularly relevant and hence aggregation poses a problem insofar that (16) does not yield an upper bound for the total risk involved, i.e. the inequality sign ($\le$) in (17) reverses ($\ge$).

4.2 An optimization example

The example below is based on Embrechts and Puccetti (2005) and the references therein. A more detailed discussion is to be found in McNeil et al. (2005).

In the language of the previous subsection, consider the loss distribution functions $F_i := F_{L_i}$, $i = 1, \dots, d$ and suppose that we are given (statistical) models or estimates for these. Further, as risk measure we take $\varrho = \mathrm{VaR}_\alpha$, for some $\alpha \in (0,1)$, typically close to 1. Since we do not assume any particular model for the interdependence between the L_i's, we do not know the joint distribution of the vector $(L_1, \dots, L_d)$ and hence we cannot calculate $\mathrm{VaR}_\alpha(\sum_{i=1}^d L_i)$. The problem then becomes: determine tight upper and lower bounds VaR_α^u, VaR_α^l so that

$$\mathrm{VaR}_\alpha^l \le \mathrm{VaR}_\alpha\Big(\sum_{i=1}^{d} L_i\Big) \le \mathrm{VaR}_\alpha^u\,. \tag{18}$$

For the calculation of these bounds, we solve an optimization problem over all joint models for $(L_1, \dots, L_d)$ keeping the marginal distribution functions $F_1, \dots, F_d$ fixed. This leads to the calculation of:

$$\begin{aligned} \mathrm{VaR}_\alpha^l &= \inf\Big\{\mathrm{VaR}_\alpha\Big(\sum_{i=1}^d L_i\Big) : L_i \sim F_i, i = 1, \dots, d\Big\}, \\ \mathrm{VaR}_\alpha^u &= \sup\Big\{\mathrm{VaR}_\alpha\Big(\sum_{i=1}^d L_i\Big) : L_i \sim F_i, i = 1, \dots, d\Big\}. \end{aligned} \tag{19}$$

When some information on the dependence between the L_i's is given, for instance expressed in terms of copulas, then one can adjust the optimization problem (19) accordingly. See Embrechts and Puccetti (2005) for details. As stated, (19) is referred to as a Fréchet problem. The mathematics used in its solution is related to the so-called mass transportation problem in measure theory; see Puccetti (2005) for a discussion of this link.

In order to illustrate the potential of the above theory, we discuss a stylized example taken from McNeil et al. (2005). From Moscadelli (2004) and further supported by deFontnouvelle (2005), we know that operational loss data often have power-like (Pareto-type) tail behavior, i.e. $F_i(x) = \mathsf{P}(L_i \le x) = 1 - x^{-\alpha_i} h_i(x)$ for some (so-called) slowly varying functions h_i and tail-index parameters α_i (i.e. $1/\kappa_i$ in Section 2.2), $i = 1, \dots, d$. Note that for $0 < \alpha_i < 1$, $\mathrm{E}(L_i) = \infty$, where for $1 < \alpha_i < 2$, $\mathrm{E}(L_i) < \infty$ but $\mathrm{Var}(L_i) = \infty$. For the further understanding of the example, the function h_i above can be neglected; readers interested in a discussion of slowly varying functions and their use in extreme value theory can consult Embrechts et al. (1997), they also briefly appeared in Section 2.3.

For the sake of simplicity, we consider $d = 8$, corresponding to the eight business lines defined in the Basel II proposal for operational risk. We assume that the marginal loss distribution functions F_i, $i = 1, \dots, 8$ are all exact Pareto $F_i(x) = \mathsf{P}(L_i \le x) = 1 - (x+1)^{-1.5}$, $x \ge 0$. Hence $\alpha_i = 1.5$, $h_i(x) \sim 1$ for $x \to \infty$, $i = 1, \dots, 8$, corresponding to a finite mean $\mathrm{E}(L_i) = 2$, infinite variance model. Under no assumption on the dependence between $L_1, \dots, L_8$, the optimization problem (18), (19) can be solved numerically. For a discussion on the sharpness of these bounds, see Embrechts and Puccetti (2005). The following results are obtained.

One first easily checks that in the comonotonic case (figures expressed in thousands and rounded):

$$\begin{aligned} \mathrm{VaR}_{0.99}\Big(\sum_{i=1}^8 L_i\Big) &= \sum_{i=1}^8 \mathrm{VaR}_{0.99}(L_i) = 0.16, \\ \mathrm{VaR}_{0.999}\Big(\sum_{i=1}^8 L_i\Big) &= \sum_{i=1}^8 \mathrm{VaR}_{0.999}(L_i) = 0.79. \end{aligned}$$

In reality however, under no specific information on the interdependence between $L_1, \dots, L_8$, one can show that $\mathrm{VaR}_\alpha\left(\sum_{i=1}^8 L_i\right)$ can reach values up to 0.41 for $\alpha = 0.99$ and for $\alpha = 0.999$ up to 1.93, more than doubling the capital charges. If however some dependence information is available then these upper bounds come down; see Embrechts and Puccetti (2005). For the sake of this example, we do not distinguish between VaR (which is just the quantile) or the mean corrected version typically used for capital allocation $\mathrm{VaR}_\alpha(L) - \mathrm{E}(L)$. The above example can be generalized to situations where the marginal Pareto distributions have different tail parameters; see Puccetti (2005).

The above example shows that, especially for heavy-tailed loss data, there is considerable uncertainty in the calculation of risk measures when no specific dependence assumption between the various risk classes can be made. Clearly, the extreme situation of *no information at all* may be far away from what information actually is available. Until now, operational risk data are too scarce in order to come up with specific dependence conditions between various operational risk classes. A viable alternative to some of the above calculations is through a combination of expert information on loss events, together with some kind of loss distribution estimation as is for instance discussed in Ebnoether et al. (2003) and the references therein.

4.3 The one loss causes ruin problem, revisited

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 very natural; see Embrechts et al. (1997). Take $X_1, \dots, X_n$ positive iid random variables with common distribution function F, denote $S_n = \sum_{k=1}^n X_k$ and $M_n = \max(X_1, \dots, X_n)$. The distribution function F is called subexponential (denoted $F \in \mathcal{S}$) if

$$\lim_{x \to \infty} \frac{\mathsf{P}(S_n > x)}{\mathsf{P}(M_n > x)} = 1 \tag{20}$$

i.e. the total loss S_n is mainly determined by one large loss M_n. Examples satisfying (20) are Pareto-type distributions, lognormal and log-gamma for instance.

A further interesting property for subexponential distributions, relevant for operational risk, uses the language of the previous sections. Suppose L_i is compound Poisson with intensity λ_i and loss distribution F_i, $i = 1, \dots, d$, so that $L_1, \dots, L_d$ are independent. One easily shows that $L = \sum_{i=1}^d L_i$ is compound Poisson with intensity $\lambda = \sum_{i=1}^d \lambda_i$ and loss distribution

$F = \sum_{i=1}^{d} \frac{\lambda_i}{\lambda} F_i$. The "one loss causes ruin" paradigm in this case translates into the fact that the loss distribution F_i with the heaviest tail determines the tail of the distribution of L. An example of such a result in the Pareto case is as follows. Suppose $1 - F_i(x) = x^{-\alpha_i} h_i(x)$ with $\alpha_1 < \alpha_2 < \cdots < \alpha_d$, then

$$\lim_{x \to \infty} \frac{\mathsf{P}(L > x)}{1 - F_1(x)} = \lambda_1, \tag{21}$$

so that $\mathsf{P}(L > x)$ for x large is mainly determined by the individual loss distribution L_1 with the heaviest tail, i.e. the smallest tail-index α_1. The above result (21) can be formulated for $F_i \in \mathcal{S}$ and more general counting random variables $N_i, i = 1, \ldots, d$. For details, see Embrechts et al. (1997), McNeil et al. (2005) and the references therein. Practitioners are well aware of this phenomenon: it is the few largest losses that cause the main concern.

5 Conclusion

As stated in the introduction, we have not attempted to review all potential approaches for the quantitative modelling of operational risk, but rather concentrated on the presentation of some of the techniques which were introduced to quantitative risk management in other publications. The references given will guide the interested reader to several of the alternative attempts available in the literature. We would like to stress that, whereas the techniques presented lend themselves ideally for most operational risk data, the same techniques have a much broader range of applications, as there are for instance the modelling of credit risk or non-life insurance data.

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.

References

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.

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.

Daley, D. and Vere-Jones, D. (1988). *An Introduction to the Theory of Point Processes*. Springer, N.Y.

deFontnouvelle, 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.

Ebnoether, S., Vanini, P., McNeil, A., and Antolinez-Fehr, P. (2003). Operational risk: A practitioner's view. *Journal of Risk*, 5(3):1–15.

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

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

Embrechts, P., Höing, A., and Juri, A. (2003b). Using copulae to bound the value-at-risk for function of dependent risks. *Finance and Stochastics*, 7(2):145–167.

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. and Puccetti, G. (2005). 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.

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.

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

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

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.

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

Neslehova, J. and Pfeifer, D. (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.

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.