Applied Probability Trust (10 October 2007)

AGGREGATING RISK ACROSS MATRIX STRUCTURED LOSS DATA: THE CASE OF OPERATIONAL RISK

PAUL EMBRECHTS,* *ETH Zurich*

GIOVANNI PUCCETTI,** *University of Firenze*

Abstract

We study the problem of evaluating the risky position involved in a matrix of random losses with some given probabilistic structure. In the Basel II regulatory setup for operational risk in banking, we analyse how interdependencies between individual loss random variables within the matrix may influence different estimates for the minimum capital charge required.

Keywords: Risk aggregation; Operational risk; Point processes; Copula methods; Dependency bounds

AMS 2000 Subject Classification: Primary 91B30

Secondary 60E15

1. Introduction and basic notation

The problem discussed in this paper concerns the measurement of risk for random losses structured in a $r \times c$ matrix. The basic language used is that of Quantitative Risk Management (QRM) as detailed in McNeil et al. (2005). Throughout the paper we will recall some of the basic notation and language from the latter so as to make the paper sufficiently self-contained for the more general reader. The mathematical problem investigated is based upon a concrete question coming from the realm of finance as discussed in the Basel Committee on Banking Supervision (2006) and summarized in Chapters 1 and 10 of McNeil et al. (2005).

The current regulatory framework for banking supervision, referred to as Basel II, allows large international banks to come up with internal models for the calculation of risk capital. Particularly for market and credit risk, a whole methodology has been worked out and is

* Postal address: Department of Mathematics, ETH Zurich, 8092 Zurich, Switzerland

** Postal address: Department of Mathematics for Decisions, University of Firenze, 50134 Firenze, Italy

applied throughout the industry. The question discussed in this paper stems from the third big class of risks falling under the internationally agreed Basel II framework: Operational Risk (OR). OR is defined as *the risk of losses resulting from inadequate or failed internal processes, people and systems, or external events. In this definition, legal risk is included. Strategic and reputational risk are excluded.* Without entering into details, the main difference with market and credit risk is that for OR we only have losses. OR concerns quality control of a financial institution. Mathematically, its modelling is very much akin to actuarial techniques mainly used in non-life insurance; see for instance McNeil et al. (2005), Chapter 10, and Panjer (2006) on these similarities. Under the so-called Loss Distribution Approach (LDA) within Basel II, financial institutions are given full freedom concerning the stochastic modelling assumptions used.

Modern QRM standardly concerns a vector of one-period profit-and-loss random variables (rvs) $\boldsymbol{X} = (X_1, \dots, X_d)'$ defined on some probability space $(\Omega, \mathfrak{F}, \mathbb{P})$ satisfying some integrability conditions. On $\boldsymbol{X}$ one defines a financial position $\psi(\boldsymbol{X})$ for some measurable function $\psi : \mathbb{R}^d \to \mathbb{R}$. A risk measure ϱ now maps $\psi(\boldsymbol{X})$ to $\varrho(\psi(\boldsymbol{X})) \in \mathbb{R}$, to be interpreted as the regulatory capital needed to be able to hold the risky position $\psi(\boldsymbol{X})$ over a predetermined fixed period. Several authors have discussed desirable properties which a risk measure ϱ has to satisfy. The interested reader can start from the seminal paper Artzner et al. (1999) where the notion of *coherent* risk measure is introduced. Textbook treatments are Delbaen (2000), McNeil et al. (2005) and Föllmer and Schied (2004). For our purposes, the axiom of *subadditivity* plays an important role: for two loss rvs L_1, L_2,

$$\varrho(L_1 + L_2) \leq \varrho(L_1) + \varrho(L_2), \tag{1}$$

but more on this later.

The standard setup for OR modelling deviates somehow from the above by organizing risk rvs not in a random vector but in a random $r \times c$ matrix $\mathcal{L} = (L_{i,j}), i = 1, \dots, r, j = 1, \dots, c$. This structure comes from the fact that, in case of OR, the Basel II framework supports the structuring of a financial institution into $r = 8$ business lines (BL) and $c = 7$ risk types (RT). For instance $L_{2,3}$ could stand for next year's total loss for BL 2 (corporate finance) and RT 3 (internal fraud). Depending on the complexity of the bank, one may deviate from this standard $56 = 8 \times 7$ structure; some banks using the LDA have less, some have more cells. The risk measure used is Value-at-Risk, $\varrho = \text{VaR}$, at the 99.9% confidence level and for a 1 year *holding* period hence corresponding to the calculation/estimation of a 1 in 1000 year loss across the OR loss matrix $\mathcal{L}$. As definition of Value-at-Risk we take the standard definition of a quantile:

$$\text{VaR}_\alpha(X) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \leq x) \geq \alpha\}. \tag{2}$$

The stochastic structure of the losses for each cell (i, j) is of the marked point process type with random intensities/severities. The basic statistical properties of OR data are discussed in the regulatory review papers Moscadelli (2004) and Dutta and Perry (2006). OR data are typically skew and very heavy-tailed. This causes problems with the axiom of subadditivity (1) in the case of $\varrho =$ VaR. In Chavez-Demoulin et al. (2006), Nešlehová et al. (2006) and Degen et al. (2007) several mathematical and statistical issues related to VaR-based risk measurement for OR are given. In this paper we assume that the marginal VaR numbers used are given and hence only concentrate on the problem of aggregation.

The basic structure of the OR matrix $\mathcal{L}$ takes the form given in Table 1. The Basel II framework allows for several levels of risk aggregation. For this purpose we introduce the following notation:

$$\begin{aligned} L_{\bullet,j} &= \sum_{i=1}^{r} L_{i,j},\ j = 1, \ldots, c, \\ L_{i,\bullet} &= \sum_{j=1}^{c} L_{i,j},\ i = 1, \ldots, r,\ \text{and} \\ L_{\bullet,\bullet} &= \sum_{i=1}^{r} \sum_{j=1}^{c} L_{i,j}. \end{aligned}$$

In principle, Basel II allows VaR estimation at any level of granularity within $\mathcal{L}$, then asks to add up the resulting VaR measures (corresponding to comonotonicity of the underlying rvs; see Section 5.1.6 in McNeil et al. (2005)), and finally allows for a reduction of this value based on diversification arguments. Concretely, most banks will model OR losses at the BL level resulting in the VaR values VaR$(L_{i,\bullet}), i = 1 \ldots, r$ and add up to obtain a risk capital estimate

$$\text{VaR}^R = \sum_{i=1}^{r} \text{VaR}\left(L_{i,\bullet}\right). \tag{3}$$

Accounting for diversification (correlation effects) would then result in a final capital estimate $(1 - \delta)\text{VaR}^R$ where $0 < \delta < 1$. Values of $\delta \simeq 0.2$ are not uncommon. Our paper tries to shed

$$
\begin{array}{cccccc}
\hline
(1,1) & (1,2) & \dots & (1,c) & \rightarrow & L_{1,\bullet} \\
\hline
(2,1) & (2,2) & \dots & (2,c) & \rightarrow & L_{2,\bullet} \\
\hline
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
\hline
(r,1) & (r,2) & \dots & (r,c) & \rightarrow & L_{r,\bullet} \\
\hline
\downarrow & \downarrow & \dots & \downarrow & \searrow & \\
L_{\bullet,1} & L_{\bullet,2} & \dots & L_{\bullet,c} & & L_{\bullet,\bullet}
\end{array}
$$

TABLE 1: Aggregate losses in matrix $\mathcal{L}$

some light on possible ranges of δ from a modelling point of view. Similarly, one can define

$$\text{VaR}^C = \sum_{j=1}^{c} \text{VaR}\left(L_{\bullet,j}\right), \text{ and} \tag{4}$$

$$\text{VaR}^+ = \sum_{i=1}^{r} \sum_{j=1}^{c} \text{VaR}\left(L_{i,j}\right). \tag{5}$$

The full bottom-up approach (5) contrasts with the top-down approach

$$\text{VaR}^T = \text{VaR}\left(L_{\bullet,\bullet}\right), \tag{6}$$

with the more realistic (3) and (4) as possible alternatives. Note that for comonotonic risks, i.e. assuming there exist increasing functions $f_{i,j}$ and a rv Z so that $\mathbb{P}(L_{i,j} = f_{i,j}(Z)) = 1$ for all i, j, we have that

$$\text{VaR}^+ = \text{VaR}^R = \text{VaR}^C = \text{VaR}^T.$$

In the absence of comonotonicity, these risk numbers will be different. In particular, we are interested in

$$\Delta = \text{VaR}^C - \text{VaR}^R, \tag{7}$$

this always at a given level α; in the case of OR, $\alpha = 0.999$. For a discussion of the difference $\text{VaR}^+ - \text{VaR}^T$, see for instance the numerous papers on the non-coherence of VaR, in particular Nešlehová et al. (2006) and Daníelsson et al. (2005).

The question on the possible values of Δ as defined in (7) was posed to us by Marco Moscadelli (private communication) and is also touched upon in Dutta and Perry (2006). As it stands, with $r \times c = 56$ and rather general assumptions on the stochastic properties of the underlying random variables, the problem is analytically extremely difficult. In the present paper we concentrate on two very specific models/examples. In the so-called *toy model* we will use some standard copula technology for severity distributions without taking the more dynamic marked point process structure of the underlying loss rvs into account. In the second, so-called *soft model*, we apply dependence modelling at this more advanced level. It is to be hoped that, through the preliminary results obtained, in the future we will be able to handle more realistic models. In the next section we recall some of the basic copula definitions.

2. Modelling of dependence: the copula approach

Recall from Section 1 that we want to investigate the properties of Δ as defined in (7). Three aspects will play an important role: the asymmetry of the loss matrix ($r \neq c$), the shape of the marginal loss distributions and the dependence between the underlying loss rvs. In order to model the latter, we use the concept of copula; see for instance Nelsen (2006) and Chapter 5 in McNeil et al. (2005). For the purpose of notation we recall some of the basic facts about copulas. A copula C is a d-dimensional distribution function (df) on $[0,1]^d$ with uniform marginals. Given a copula C and d univariate marginals $F_1, \ldots, F_d$, one can always define a df F on $\mathbb{R}^d$ having these marginals by

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)),\ x_1, \ldots, x_d \in \mathbb{R}. \tag{8}$$

Sklar's theorem (see Nelsen (2006)) states conversely that we can always find a copula C coupling the marginal dfs of a fixed joint df F trough the above expression (8). For continuous marginal dfs, this copula is unique. For discrete dfs, one has to be somewhat careful; see Genest and Nešlehová (2007) for a discussion of the discrete case. Any copula C satisfies the Fréchet bounds

$$\max\left\{\sum_{i=1}^{d} u_i - d + 1, 0\right\} \leq C(u_1, \ldots, u_d) \leq \min\{u_1, \ldots, u_d\},\ u_1, \ldots, u_d \in [0,1].$$

In the following, we denote the upper Fréchet bound by $M(u_1, \ldots, u_d)$. M is the so-called *comonotonic* copula already mentioned in Section 1. See Dhaene et al. (2002) for a detailed discussion of the concept of comonotonicity within QRM. In order to model some dependence scenarios between OR losses, we will use, besides the comonotonic copula M and the independence copula $\Pi(u_1, \ldots, u_d) = \prod_{i=1}^{d} u_i,\ u_1, \ldots, u_d \in [0,1]$, the following copula families:

(i) The *Gumbel* copula with parameter $\theta \geq 1$,

$$C_\theta^{Gu}(u_1,\ldots,u_d) = \exp\left(-\left[(-\ln u_1)^\theta + \cdots + (-\ln u_d)^\theta\right]^{1/\theta}\right),$$

The Gumbel copula interpolates between independence ($C_1^{Gu} = \Pi$) and comonotonic dependence ($C_{+\infty}^{Gu} = M$). The parameter θ can easily be calibrated using Kendall's tau $\tau = 1 - \theta^{-1}$ and exhibits upper tail dependence $\lambda_u(C_\theta^{Gu}) = 2 - 2^{1/\theta}$ as explained in McNeil et al. (2005, page 222). Also note that the lower dimensional coordinate projections of the Gumbel copula are again Gumbel, a property which will come in handy later. Algorithm 5.48(b) in McNeil et al. (2005) explains how to generate data from C_θ^{Gu}.

(ii) The *Gaussian* copula with correlation parameter $-1 \leq \rho \leq 1$,

$$C_\rho^{Ga}(u_1,\ldots,u_d) = \mathbf{\Phi}_\rho\left(\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_d)\right). \tag{9}$$

Here $\mathbf{\Phi}_\rho$ denotes the joint distribution of a zero-mean Gaussian random vector with equicorrelation matrix Σ, with ones on the main diagonal and off-diagonal elements equal to ρ. Φ^{-1} is the quantile function of a standard normal rv. Analogously to the Gumbel, the Gaussian copula interpolates between independence ($C_0^{Ga} = \Pi$) and the comonotonic limiting case ($C_1^{Ga} = M$). In contrast however to C_θ^{Gu}, C_ρ^{Ga} for $\rho < 1$ is asymptotically tail independent, $\lambda_u(C_\rho^{Ga}) = 0$; see McNeil et al. (2005). Also the Gaussian copula family is closed under coordinate projections and simulation of Gaussian copula data is straightforward.

3. A toy model for OR matrix data

In order to understand some of the main drivers determining the properties of Δ in (7), we start with an easy, stylized model for the OR losses in a data matrix $\mathcal{L}$. In this, so-called *toy model*, we keep non-squaredness of the matrix by putting $r = 2, c = 3$; extensions to more realistic dimensions (like $r = 8, c = 7$) is a matter of computational power.

For each cell (i,j) of $\mathcal{L}$, we assume a Pareto model with a tail parameter α in the range of observed values as reported in the Quantitative Impact Studies of Moscadelli (2004) and Dutta and Perry (2006), i.e. we take $\alpha \in [1,4]$. Moscadelli (2004) even reports infinite mean models ($\alpha < 1$), whereas Dutta and Perry (2006) typically find values up to 5. Concretely, we assume for the toy model that for $i = 1, 2, j = 1, 2, 3$,

$$F_{L_{i,j}}(x) = \mathbb{P}(L_{i,j} \leq x) = 1 - (1 + x)^{-\alpha_{i,j}}, x \geq 0.$$

We couple the marginal loss dfs $F_{L_{i,j}}$ via a Gumbel copula with parameter $\theta \geq 1$ so that

$$\mathbb{P}\left(\cap_{i,j}\{L_{i,j} \leq x_{i,j}\}\right) = C_\theta^{Gu}(F_{L_{1,1}}(x_{1,1}), F_{L_{1,2}}(x_{1,2}), F_{L_{1,3}}(x_{1,3}), F_{L_{2,1}}(x_{2,1}), F_{L_{2,2}}(x_{2,2}), F_{L_{2,3}}(x_{2,3})).$$

In order to incorporate one further degree of freedom into our analysis, we differentiate between a homogeneous (T1) and a non-homogeneous model (T2), where

- Model T1: $(\alpha_{i,j}) = \begin{pmatrix} 4 & 4 & 4 \\ 4 & 4 & 4 \end{pmatrix}$.
- Model T2: $(\alpha_{i,j}) = \begin{pmatrix} 1.25 & 2 & 2.75 \\ 2 & 2.75 & 3.5 \end{pmatrix}$.

An explicit analysis of Δ is possible *only* within the homogeneous model T1, for which VaR^C and VaR^R can be obtained analytically. Within model T2, only VaR^C can be obtained analytically, and this because this latter estimate is based on the VaRs of the two-dimensional submodels $L_{\bullet,j}, j = 1, 2, 3$. In fact, estimation of the quantiles for the sum of d rvs coupled by a generic copula is difficult when $d = 3$ *and* the model is not homogeneous, or when $d \geq 4$ under any dependence model. An exception is represented by the comonotonic scenario $\theta = +\infty$, where quantile estimates are always straightforward due to VaR additivity in this case.

Therefore, apart from the above mentioned cases, we resort in our first analysis to using simulation. For this, we generate 10^6 realizations using the dependence structures M, Π and C_θ^{Gu} for some θ values. For the calculation of the resulting VaRs, we use Extreme Value Theory (EVT) methodology in its Peaks Over Threshold (POT) form as explained in McNeil et al. (2005, Sect. 7.2) and Embrechts et al. (1997, Sect. 6.5); the R version of the library QRMlib for this analysis can be downloaded from the website of Alexander J. McNeil `http://www.ma.hw.ac.uk/~mcneil/book/QRMlib.html`. See also that webpage for full statistical details of the procedure. EVT is widely used within the realm of OR modelling; see Moscadelli (2004). For some criticism however, see Dutta and Perry (2006) and Section 5 below. In the case of the exact Pareto models above, EVT estimation performs very well. In

Model T1

θ	VaR^C	VaR^R	VaR^T	Δ
1.00	18.31	14.46	9.96	$\mathbf{3.85}$
1.10	21.32	19.38	17.70	$\mathbf{1.94}$
1.25	23.71	22.62	21.93	$\mathbf{1.09}$
$+\infty$	27.74	27.74	27.74	$\mathbf{0}$

Model T2

θ	VaR^C	VaR^R	VaR^T	Δ
1.00	299.28	289.71	257.17	$\mathbf{9.57}$
1.10	309.47	301.03	279.98	$\mathbf{8.44}$
1.25	318.96	315.51	302.68	$\mathbf{3.45}$
$+\infty$	340.29	340.29	340.29	$\mathbf{0}$

TABLE 2: Estimated Value-at-Risk values for models T1 and T2, together with the difference $\Delta = \text{VaR}^C - \text{VaR}^R$.

Table 2 we summarize for the models T1, T2 the various estimates for the VaRs defined in Section 1: VaR^C,VaR^R,VaR^T as well as the difference $\Delta = \text{VaR}^C - \text{VaR}^R$. Recall that all VaRs are calculated at the 99.9% confidence level.

From Table 2 we can already derive some preliminary facts. Note that under model T1 the only asymmetry comes from the fact that $r \neq c$. If we take VaR^T as a proxy for the true matrix VaR, then we see that marginal aggregation under the T1 assumption always overestimates the risk and that this overestimation is monotone in the strength of the interdependence between the losses. Also, Δ is always positive and again monotone in θ: it decreases from a maximal value given independence to 0 for comonotonicty. Similar results hold for the inhomogeneous loss matrix $\mathcal{L}$ from model T2.

In order to investigate the consequences of the Basel II guidelines for LDA-based risk capital aggregation across the OR loss matrix $\mathcal{L}$, in Table 3 we include upper and lower bounds on the Value-at-Risk for the total aggregated data, $\text{VaR}^T = \text{VaR}(L_{\bullet,\bullet})$, together with similar bounds for the row-/columnwise aggregated risks. The upper bounds are calculated using the mass transportation techniques developed in Embrechts and Puccetti (2006). The lower bounds are calculated *analytically* via expression (13) in Denuit et al. (1999). It is useful to remark that the computation of upper, respectively lower bounds on the Value-at-Risk of a df is equivalent to the calculation of lower, respectively, upper bounds on the df itself. Moreover, note that all bounds are calculated only using the marginal dfs of the loss rvs and hold across all interdependence assumptions. Such bounds are useful, especially in OR practice, as so far no widely agreed models for dependence between OR loss data have been brought forward.

Model T1

risk	LB	$\theta = 1$	$\theta = 1.1$	$\theta = 1.25$	$\theta = +\infty$	UB
$L_{\bullet,\bullet}$	4.62	9.96	17.70	21.93	27.74	38.51
$L_{\bullet,1}, L_{\bullet,2}, L_{\bullet,3}$	$\mathbf{4.62}$	6.10	7.11	7.90	9.25	$\mathbf{11.37}$
$L_{1,\bullet}, L_{2,\bullet}$	4.62	7.23	9.69	11.31	13.87	18.44

Model T2

risk	LB	$\theta = 1$	$\theta = 1.1$	$\theta = 1.25$	$\theta = +\infty$	UB
$L_{\bullet,\bullet}$	250.19	257.17	279.98	302.68	340.29	538.62
$L_{\bullet,1}$	$\mathbf{250.19}$	254.41	260.60	266.57	280.81	$\mathbf{365.71}$
$L_{\bullet,2}$	$\mathbf{30.62}$	32.42	34.94	37.21	41.95	$\mathbf{54.85}$
$L_{\bullet,3}$	$\mathbf{11.33}$	12.45	13.93	15.18	17.53	$\mathbf{22.13}$
$L_{1,\bullet}$	250.19	255.79	264.63	274.71	292.14	403.61
$L_{2,\bullet}$	30.62	33.92	36.40	40.80	48.15	67.75

TABLE 3: Estimated VaRs for row- and columnwise aggregated risks in model T1 and T2 using a Gumbel parameter $\theta \geq 1$. Lower (LB) and upper (UB) bounds are given; bold face values indicate sharpness of these bounds.

As an example, consider row 2 in the model T1 part of Table 3. For each $j = 1, 2, 3$, VaR$(L_{\bullet,j}) \in [4.62, 11.37]$ where these bounds are sharp, i.e. are attainable for a given dependence structure (copula). The gap between the upper bound UB=11.37 and the comonotonicity value ($\theta = +\infty$) 9.25 corresponds to models where Value-at-Risk is not coherent. The upper bound 38.51 for the totally aggregated loss is not necessarily sharp, similarly the 18.44 value for rowwise aggregated losses. The reason for this is that sharpness of all the bounds reported here may fail for $d \geq 3$. For Basel II practice, the last two lines $(L_{1,\bullet}, L_{2,\bullet})$ in the model T2 part of Table 3 are relevant. Recall that in $L_{i,\bullet}, i = 1, 2$, the loss data are first aggregated rowwise, after which the 99.9% VaRs are calculated. Note that the range (LB,UB) reported only reflects numerical bounds and *not* statistical uncertainty; of course, the latter further compounds the calculations in practice. As already stated, for the purpose of this paper, we neglect this important issue. Take for instance the $\theta = 1.1$ column, corresponding to about 10% Kendall τ rank correlation. The business line VaRs are calculated as $\text{VaR}(L_{1,\bullet}) = 264.63$ and $\text{VaR}(L_{2,\bullet}) = 36.40$, assuming the same Gumbel dependence for the individual losses. The LB and UB values give non-sharp numerical bounds for these values. Given that the bank uses the Pareto parameter values from model T2 and assumes a Gumbel dependence model with $\theta = 1.1$ for its LDA, then the Basel II guidelines require the bank to first report $\text{VaR}(L_{1,\bullet}) + \text{VaR}(L_{2,\bullet}) = 264.63 + 36.40 = 301.03$. This latter value can also be found in the second row of the Model T2 part of Table 2. The values for $\text{VaR}(L_{\bullet,\bullet})$ yield an indication what would be possible with respect to so-called diversification effects (LB=250.19) but also the upper bound UB=538.62 shows how far non-coherence of VaR may influence the capital charge *across* all dependence models. For the specific Gumbel dependence $\theta = 1.1$, we have subadditivity both column- as well as rowwise:

$$\text{VaR}^T \leq \min\{\text{VaR}^C, \text{VaR}^R\}. \tag{10}$$

While still far away from a realistic OR data matrix $\mathcal{L}$, the above discussion shows the full complexity of the capital charge issue of OR within an LDA.

In the next section, we add a first level of model complexity by separating the loss rvs for each cell in intensity and severity components.

4. A soft model for OR matrix data

We base our intensity-severity *soft model* on the point process model developed in Pfeifer and Nešlehová (2004); see also Chavez-Demoulin et al. (2006) for a summary in an OR context. Alternatively, we could have used the common Poisson shock model as presented in Lindskog and McNeil (2003). As in the previous section, we restrict the calculations to an $r = 2, c = 3$ OR loss matrix $\mathcal{L}$.

Assume that the yearly aggregated OR losses for business line i and risk type j (i.e. cell

(i, j) in $\mathcal{L}$) are given by

$$L_{i,j} = \sum_{k=0}^{N_{i,j}} X_{i,j}^k,\ i = 1, 2,\ j = 1, 2, 3. \tag{11}$$

Here $X_{i,j}^k$, with $X_{i,j}^0 \equiv 0$, represents the k-th loss in cell (i, j) with $N_{i,j}$ the random total number of losses over a one year period, say, in cell (i, j). So far, standard literature like Dutta and Perry (2006) assumes independence between as well as within the $(N_{i,j})_{i,j}$ and the loss occurrences $(X_{i,j}^k)_k$. Models like (11) belong to the realm of classical insurance risk theory as for instance discussed in Panjer (2006) in an OR context.

In the soft model below we shall make both the frequency variables $N_{i,j}$ as well as the severity variables $X_{i,j}^k$ dependent. Throughout the OR literature, the homogeneous Poisson model has become somewhat of a standard, hence we assume for the soft model that, for all i, j, $N_{i,j} \overset{\mathrm{d}}{=}$ Pois($\lambda = 20$); the homogeneity across the intensities is taken for numerical convenience. For a detailed discussion of such models used within the LDA of a large international bank, see Aue and Kalkbrenner (2006). The six discrete rvs $N_{i,j}, i = 1, 2, j = 1, 2, 3$ are coupled to a joint model using a copula $C^f : [0, 1]^6 \to [0, 1]$. At this point, we would like to recall the problem of copula modelling for discrete marginals as for instance discussed in Genest and Nešlehová (2007). The Poisson assumption above, combined with the copula C^f yields a joint model for the cell frequencies. Within each cell (i, j) we assume that the individual losses $(X_{i,j}^k), k = 1, \ldots, N_{i,j}$ are iid with common df $F_{X_{i,j}}$, and are independent of $N_{i,j}$. The dependence between the k-th severities of the different cells is governed by a copula $C^s : [0, 1]^6 \to [0, 1]$. As a consequence, given the compound loss structure in (11), the stochastic structure of (8) is determined by the joint dfs

$$C^f(F_{N_{1,1}}, F_{N_{1,2}}, F_{N_{1,3}}, F_{N_{2,1}}, F_{N_{2,2}}, F_{N_{2,3}})$$

for cell frequencies and

$$C^s(F_{X_{1,1}}, F_{X_{1,2}}, F_{X_{1,3}}, F_{X_{2,1}}, F_{X_{2,2}}, F_{X_{2,3}})$$

for the k-th cell severities. Provided that we do not consider time occurrences of OR losses, this construction method is similar to the one described in Section 4.3.2 in Chavez-Demoulin et al. (2006). A full mathematical description of this model, using the language of marked point processes, is to be found in Section 4 of Pfeifer and Nešlehová (2004).Throughout the discussion below, subscript f, respectively s, stands for frequency, respectively severity.

The consequences of the above assumptions for the agrregation of OR risk measures are analysed for the following models (recall that for all i, j, $F_{N_{i,j}}$ = Pois(20)):

- Model S1: Gumbel & Pareto, $C^f = C^{Gu}_{\theta_f}, C^s = C^{Gu}_{\theta_s}$, $F_{X_{i,j}}$ = Pareto(4).
- Model S2: Gumbel & Lognormal, $C^f = C^{Gu}_{\theta_f}, C^s = C^{Gu}_{\theta_s}$, $F_{X_{i,j}}$ = LN(μ, σ^2) (μ and σ^2 are chosen so as to match the first two moments of Pareto(4)).
- Model S3: Gaussian & Pareto, $C^f = C^{Ga}_{\rho_f}, C^s = C^{Ga}_{\rho_s}$, $F_{X_{i,j}}$ = Pareto(4).

The choices of the claim dfs (Pareto, LN) reflect the semi- to heavy-tailedness of typical OR losses; the dependence structures (C^{Gu}, C^{Ga}) allow for asymptotic tail dependence versus independence models. At this point, there is no further compelling reason behind the choice of copulas beyond their illustrative character.

With respect to the models S1, S2 and S3, analytic calculations are out of the question so that we have to resort to simulation. For this, we generate 10^6 realizations of a bank's one year OR loss experience over the 16 severity/frequence scenarios given. As for the toy model, 99.9% VaRs are calculated using EVT.

In Table 4 we have summarized the results for model S1 giving VaR^C, VaR^R, VaR^T and Δ as defined in Section 1 ((4), (3), (6) and (7)). Recall that the original interest shown by banking regulators in the problem treated in this paper concerns the properties of Δ, the difference between columnwise (risk type) VaR aggregation and rowwise (business type) VaR aggregation. From Table 4 we see that the largest Δ value is obtained under full independence ($\theta_s = \theta_f = 1$) with $\Delta = 0$ in the comonotonic case. In all directions of the table, Δ decreases with increasing dependence. Moreover, interdependence among severities seems to play a bigger role in reducing Δ with respect to interdependence among frequencies. The situation is consistent across all three models (Tables 4, 5 and 6). Comparing Tables 4 and 5 we note that the light-tailed LN yields bigger VaR estimates; it is too early to draw firm conclusions from this, as the confidence level (99.9%) also has to be taken into account and we may not be far enough in the tail of the aggregate dfs as to see full heaviness of the Pareto tail.

In the transition from Table 4 to Table 6, Gaussian interdependence causes an overall reduction of all quantile estimates, with the exception of the four *corner scenarios* in which the models S1 and S3 are identical. This fact may be caused by the different upper tail dependence behavior of the Gaussian copula with respect to the Gumbel copula. We also want to remark that the linear correlation used in Table 6 yield Kendall τ values similar to those attained in

Table 4 and 5. We are aware that Kendall τ rank correlation has to be treated with care in case of frequencies; see Genest and Nešlehová (2007) on this. Analogously to the soft models analysed in Section 3, we have subadditivity of VaR as indicated in (10). The inequality in (10) may be reversed with the transition to infinite mean loss distributions; see Nešlehová et al. (2006). Finally note that we cannot provide a version of Table 3 for our soft models, since loss aggregated dfs in each cell of $\mathcal{L}$ are not known analytically; see Section 5.1 in Embrechts and Puccetti (2006) on this.

Model S1	independent severities ($\theta_s = 1$)	dependent severities ($\theta_s = 1.1$)	dependent severities ($\theta_s = 1.25$)	comonotonic severities ($\theta_s = +\infty$)
	89.43	96.18	102.06	111.39
independent	78.08	89.92	97.52	109.10
frequencies	64.91	83.38	93.05	107.76
($\theta_f = 1$)	**11.35**	**6.26**	**4.54**	**2.29**
	90.39	97.92	103.53	111.78
dependent	80.18	91.90	99.24	110.10
frequencies	70.77	86.29	95.43	108.90
($\theta_f = 1.1$)	**10.21**	**6.02**	**4.29**	**1.68**
	91.17	98.70	104.31	112.59
dependent	82.18	92.76	100.80	111.66
frequencies	73.60	87.65	97.07	110.75
($\theta_f = 1.25$)	**8.99**	**5.94**	**3.51**	**0.93**
	93.42	101.58	106.65	117.90
comonotonic	85.36	96.96	103.80	117.90
frequencies	77.21	92.94	101.20	117.90
($\theta_f = +\infty$)	**8.06**	**4.62**	**2.85**	**0.00**

TABLE 4: Values for VaR^C (first row in each cell), VaR^R (second), VaR^T (third) and Δ (fourth, boldfaced) under 16 different dependence scenarios in model S1.

Model S2	independent severities ($\theta_s = 1$)	dependent severities ($\theta_s = 1.1$)	dependent severities ($\theta_s = 1.25$)	comonotonic severities ($\theta_s = +\infty$)
	106.29	118.08	125.58	140.49
independent	91.10	108.80	120.24	138.32
frequencies	72.64	101.07	114.85	135.92
($\theta_f = 1$)	$\mathbf{15.19}$	$\mathbf{9.28}$	$\mathbf{5.34}$	$\mathbf{2.17}$
	107.82	118.77	127.05	141.36
dependent	92.72	110.18	121.92	139.50
frequencies	76.45	102.90	116.75	138.09
($\theta_f = 1.1$)	$\mathbf{15.10}$	$\mathbf{8.59}$	$\mathbf{5.13}$	$\mathbf{1.86}$
	108.36	119.73	127.59	143.10
dependent	93.76	111.26	122.56	141.72
frequencies	78.53	103.36	117.59	140.38
($\theta_f = 1.25$)	$\mathbf{14.60}$	$\mathbf{8.47}$	$\mathbf{5.03}$	$\mathbf{1.38}$
	110.61	122.67	131.67	146.88
comonotonic	96.94	114.86	127.82	146.88
frequencies	82.69	108.54	124.02	146.88
($\theta_f = +\infty$)	$\mathbf{13.67}$	$\mathbf{7.81}$	$\mathbf{3.85}$	$\mathbf{0.00}$

TABLE 5: The same as Table 4 for model S2.

Model S3	independent severities ($\rho_s = 0$)	dependent severities ($\rho_s = 0.142$)	dependent severities ($\rho_s = 0.309$)	comonotonic severities ($\rho_s = 1$)
	89.43	89.76	92.37	111.39
independent	78.08	79.84	83.24	109.10
frequencies	64.91	67.96	72.77	107.76
($\rho_f = 0$)	$\mathbf{11.35}$	$\mathbf{9.92}$	$\mathbf{9.13}$	$\mathbf{2.29}$
	89.49	90.84	92.91	112.35
dependent	78.96	81.02	83.98	110.44
frequencies	66.73	69.97	74.01	108.74
($\rho_f = 0.142$)	$\mathbf{10.53}$	$\mathbf{9.82}$	$\mathbf{8.93}$	$\mathbf{1.91}$
	90.12	91.68	93.69	112.68
dependent	80.10	82.52	85.32	111.20
frequencies	68.83	72.02	76.41	109.57
($\rho_f = 0.309$)	$\mathbf{10.02}$	$\mathbf{9.16}$	$\mathbf{8.37}$	$\mathbf{1.48}$
	93.42	95.46	97.35	117.90
comonotonic	85.36	87.64	91.02	117.90
frequencies	77.21	80.39	84.48	117.90
($\rho_f = 1$)	$\mathbf{8.06}$	$\mathbf{7.82}$	$\mathbf{6.33}$	$\mathbf{0.00}$

TABLE 6: The same as Table 4 for model S3.

5. Numerical calculation and statistical estimation

In the analysis of the toy and the soft models, the large number of simulated data allowed us to avoid estimation issues in detail. In this section, we briefly address some statistical issues one may face in real OR applications.

The quantile estimates of the previous sections were based on EVT. We are aware that there exist different methods to obtain an estimate of the quantile of an aggregate distribution when the latter is not known analytically. Especially when the random variables involved are dependent, Monte Carlo techniques are often called for.

As *crude* Monte Carlo estimates are widely used throughout the OR literature, we want to stress that such VaR estimates are obtained as the quantile of the simulated empirical distribution, and this without a confidence interval. This latter methodology may not work, or produce a relatively large confidence interval for the resulting quantile when simulated samples show heavy-tailedness and VaR is computed at a probability level $\alpha \geq 0.999$.

Even more sophisticated techniques such as *importance sampling* and *bootstrap methods* may encounter problems in the case of a heavy-tailed distribution. While importance sampling with long-tailed data is still in its infancy, see for instance Asmussen et al. (2000), the bootstrap method does not manage to extrapolate into a heavy tail, since the empirical df terminates at the largest sample point. In order to calculate the probability of rare events, we prefer to use the semi-parametric techniques provided by EVT. For a more in depth discussion on the difference between empirical tail estimation versus EVT-based estimation, see de Haan and Ferreira (2006).

Of course, this choice does not come without a cost. The statistical reliability of the POT approach is very sensitive to the choice of the threshold u beyond which a GPD distribution is fitted. In fact, a critical bias-variance trade-off arises at this point and several methods for choosing u in the POT analysis are available in the literature; see Beirlant et al. (2004) for a review. However, in our 10^6 data-world, 0.999-VaR estimates proved to be very robust against threshold selection, their 99% confidence intervals being so narrow that this problem can be bypassed for our application. In conclusion, the POT-GPD approach yields an excellent fit to the tail of the distributions under study.

In real practice, poor availability of data and dependence arising between observations (non-stationarity, clustering, trends, etc...) make EVT much more problematic. An in-depth discussion of all the remaining cautions to be exercised while performing an EVT analysis can be found in Embrechts et al. (1997, Section 6.5). More fallacies and issues underlying the AMA-LDA modelling of OR are included in Nešlehová et al. (2006). A good alternative, which seems to produce a good fit in the tail regions of OR losses, is the fitting of a *g-and-h* distribution; see Dutta and Perry (2006). On the interplay between EVT-based and g-and-h estimation for OR losses, see Degen et al. (2007) and Jobst (2007).

We also ignore the problem of testing statistically the positivity and monotonicity of Δ as a function of the dependence parameters, which seems to be analytically a very difficult problem. Moreover, we provide the same quantile estimates for different data sets sampled from the same df. As an example, take the four corner scenarios in Table 4 (model S1) and Table 6 (model S3). In these cases, the Gumbel, respectively Gaussian copula assumption leads to the same probabilistic setup within model S1, respectively S3. In line with this fact, we provide the same quantile estimates for these scenarios, getting rid of statistical uncertainty.

As already stated in the beginning of the paper, we do not treat the many remaining statistical problems (reporting bias, limited collection periods, cut-off values, etc. . .) typically present in the analysis of OR data. For these issues, we refer to de Fontnouvelle et al. (2005).

Finally some comments on more realistic OR data matrices $\mathcal{L}$. Simulating data for a 8×7-dimensional matrix is a matter of computational time and memory; and this both for the toy and the soft models presented above. When using a symmetric model with a Gumbel or Gaussian copula coupling homogeneous marginals, EVT analysis can be reduced to considering the first row and, respectively, the first column of $\mathcal{L}$, since the $(L_{i,\bullet})_i$ and, respectively, the $(L_{\bullet,j})_j$ are identically distributed. The inclusion of non-homogeneous marginals, which is typically required within an OR context, requires much more effort. In this latter case, one should estimate $7 + 8 = 15$ quantiles via EVT, and this for all the dependence scenarios in the model. In the soft model, where two dependence parameters are allowed to vary, this calls for an automated threshold selection procedure. Even with 10^6 simulated data, the threshold selection methods reviewed in Beirlant et al. (2004) did not prove to be reliable across the different dependence scenarios. No doubt more research on this is needed.

6. Conclusion and future work

In this paper we have presented a methodology for the analysis of risk across a non-symmetric matrix of loss data. This study was motivated by specific questions related to the 8×7 loss matrix for operational risk. At present it is too early to have concrete probabilistic assumptions for the dfs of OR loss severities and/or frequencies, as indeed for the various interdependence structures. One conclusion however is clear: aggregation at will may lead to regulatory arbitrage. Some clear guidelines on how to aggregate risk measures (VaRs) within the LDA (Loss Distribution Approach) are no doubt called for, for instance by only allowing data aggregation at business line level. For the latter, recall however that aggregation of loss data across risk type may lead to non-homogeneous data for which VaR estimation at high confidence levels ($\geq 99.9\%$) is difficult; see Nešlehová et al. (2006) on this. For our analysis, we have considered models with at least three moments finite; the case of infinite mean data, as reported in Moscadelli (2004) complicates matters further. We will return to some of these issues in future publications.

References

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999). Coherent measures of risk. *Math. Finance 9*(3), 203–228.

Asmussen, S., K. Binswanger, and B. Højgaard (2000). Rare events simulation for heavy-tailed distributions. *Bernoulli 6*(2), 303–322.

Aue, F. and M. Kalkbrenner (2006). LDA at work: Deutsche Bank's approach to quantifying operational risk. *J. Operational Risk 1*(4), 49–93.

Basel Committee on Banking Supervision (2006). *International Convergence of Capital Measurement and Capital Standards*. Basel: Bank for International Settlements.

Beirlant, J., Y. Goegebeur, J. Teugels, and J. Segers (2004). *Statistics of Extremes*. Chichester: John Wiley & Sons.

Chavez-Demoulin, V., P. Embrechts, and J. Nešlehová (2006). Quantitative models for operational risk: extremes, dependence and aggregation. *Journal of Banking and Finance 30*, 2635–2658.

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

de Fontnouvelle, P., V. DeJesus-Rueff, J. Jordan, and E. Rosengren (2005). Capital and risk: new evidence on implications of large operational losses. Working Paper 03-5, Federal Reserve Bank of Boston.

de Haan, L. and A. Ferreira (2006). *Extreme Value Theory. An Introduction.* Berlin: Springer.

Degen, M., P. Embrechts, and D. D. Lambrigger (2007). The quantitative modeling of operational risk: between g-and-h and EVT. Astin Bull., to appear.

Delbaen, F. (2000). *Coherent risk measures*. Cattedra Galileiana. Classe di Scienze, Pisa: Scuola Normale Superiore.

Denuit, M., C. Genest, and É. Marceau (1999). Stochastic bounds on sums of dependent risks. *Insurance Math. Econom. 25*(1), 85–104.

Dhaene, J., M. Denuit, M. J. Goovaerts, R. Kaas, and D. Vyncke (2002). The concept of comonotonicity in actuarial science and finance: theory. *Insurance Math. Econom. 31*(1), 3–33.

Dutta, K. and J. Perry (2006). A tale of tails: an empirical analysis of loss distribution models for estimating operational risk capital. Federal Reserve Bank of Boston, Working Paper No. 6-13.

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

Embrechts, P. and G. Puccetti (2006). Aggregating risk capital, with an application to operational risk. *Geneva Risk Insur. Rev. 31*(2), 71–90.

Föllmer, H. and A. Schied (2004). *Stochastic Finance*. Berlin: Walter de Gruyter. Second edition.

Genest, C. and J. Nešlehová (2007). A primer on copulas for count data. *Astin Bull., to appear*.

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

Lindskog, F. and A. J. McNeil (2003). Common Poisson shock models: applications to insurance and credit risk modelling. *Astin Bull. 33*(2), 209–238.

McNeil, A. J., R. Frey, and P. Embrechts (2005). *Quantitative risk management*. Princeton, NJ: Princeton University Press.

Moscadelli, M. (2004). The modelling of operational risk: experience with the analysis of the data collected by the Basel Committee. Temi di discussione, Banca d'Italia, URL:http://www.bancaditalia.it/ricerca/consultazioni/temidi/td04/td517/td_517/tema_517.pdf.

Nelsen, R. B. (2006). *An Introduction to Copulas*. New York: Springer-Verlag. Second Edition.

Nešlehová, J., P. Embrechts, and V. Chavez-Demoulin (2006). Infinite-mean models and the LDA for operational risk. *J. Operational Risk 1*(1), 3–25.

Panjer, H. H. (2006). *Operational risk*. Hoboken, NJ: John Wiley & Sons.

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