

# Loss Distribution Approach for operational risk\*

A. Frachot, P. Georges & T. Roncalli†

Groupe de Recherche Opérationnelle, Crédit Lyonnais, France

First version: March 30, 2001

This version: April 25, 2001‡

## Abstract

In this paper, we explore the Loss Distribution Approach (**LDA**) for computing the capital charge of a bank for operational risk where **LDA** refers to statistical/actuarial methods for modelling the loss distribution. In this framework, the capital charge is calculated using a *Value-at-Risk* measure. In the first part of the paper, we give a detailed description of the **LDA** implementation and we explain how it could be used for economic capital allocation. In the second part of the paper, we compare **LDA** with the Internal Measurement Approach (**IMA**) proposed by the Basel Committee on Banking Supervision to calculate regulatory capital for operational risk. **LDA** and **IMA** are bottom-up internal measurement models which are apparently different. Nevertheless, we could map **LDA** into **IMA** and give then some justifications about the choice done by regulators to define **IMA**. Finally, we provide alternative ways of mapping both methods together.

## 1 Introduction

For financial institutions, risk has several components: credit risk, market risk, other types of risk such as operational risk (Figure 1). Under the 1988 Accord, the Basel Committee on Banking Supervision recognises that the capital charge related to credit risk implicitly covers other risks. Reflecting that risks other than credit and market risks can be substantial, operational risk are now explicitly concerned by the New Basel Capital Accord. A recent survey issued by the Risk Management Group suggests that economic capital allocation for operational risk ranges between 15-25% for the majority of banks.

As a result, there is a growing pressure from supervision authorities on the management of operational risk by financial institutions. A common industry definition of the scope of operational risk is as follows: “*the risk of direct or indirect loss resulting from inadequate or failed internal processes, people and systems or from external events*”. If legal risk is generally included in this definition, reputational and strategic risks are not. This definition focuses on causes of loss, called event type, but do not precise their effects (loss type), although both event type and loss type should be identified when recording loss data. Since event risks can be identified objectively in a consistent manner accross banks, the Committee believes that this is appropriate for both risk management and measurement. The operational risk is now receiving and will receive the same regulatory treatment imposed on credit and market risks. As for these two types of risk, the regulatory treatment is now at a stage of demanding standard computations of the unexpected loss, but it appears a commitment to turn in a near future to an internal-based model. As a result, banks would be allowed to build an internal model based on a mix of internal/external data and on an in-house methodology.

---

\*We gratefully thank Maxime Pennequin, the operational risk manager, Fabienne Bieber who is in charge of the operational risk measurement in the Capital Allocation team as well as the Operational Risk team at Crédit Lyonnais for stimulating discussions. We thank Nicolas Baud for mathematical assistance on order statistics, Gaël Riboulet for the alternate proof in Footnote 24 and Anne Chartrain and Pierre Martineu for their comments and suggestions. All remaining errors are ours.

† *Corresponding author*: Groupe de Recherche Opérationnelle, Bercy-Expo — Immeuble Bercy Sud — 4<sup>ème</sup> étage, 90 quai de Bercy — 75613 Paris Cedex 12 — France; *E-mail address*: thierry.roncalli@creditlyonnais.fr

‡The last version of this paper can be downloaded at the web page <http://gro.creditlyonnais.fr>.

## Loss Distribution Approach for operational risk

---

For operational risk capital charge, an evolutionary framework of four stages is proposed. The first one, also called the Basic Indicator Approach (**BIA**), is the most straightforward approach. The required capital is determined by multiplying a financial indicator, such as gross income, by a fixed percentage (called the ‘alpha’ factor). The Standardised Approach (**SA**) differs from the latter in that banks would divide their activities into a number of standardised business units and business lines. Within each business line, the capital charge is calculated by multiplying an indicator, such as gross income or asset size of the business line, by a fixed percentage (called the ‘beta’ factor). The total capital charge will be the simple summation of the capital required across each of the business lines. In both cases, a **possible** calibration mechanism for ‘alpha’ and ‘beta’ parameters **would** be based on 20% of current regulatory capital. The Internal Measurement Approach (**IMA**) provides banks to use their internal loss data as inputs for a capital calculation but in a way given by supervisors. Operational risk is categorised according to a matrix of business lines and operational risk types, which would be standardised by supervisors. The required capital within each business line/loss type combination will be calculated by multiplying the expected loss by a fixed percentage (called the ‘gamma’ factor). Expected loss is computed as the product of an exposure indicator, standardised by supervisors as a proxy for the amount of risk of each business line/loss type combination, and two terms based on internal data: the probability of loss event and a parameter representing the loss given that event. Since the gamma factor is computed on a industry based distribution, it will be possible to adjust capital charge by a risk profile index, which reflects the bank’s specific risk profile compared to industry. The total capital charge will be the simple sum of the required capital across each of business line and risk type combinations. The most sophisticated approach, which this paper will focus on, is the Loss Distribution Approach (**LDA**). Under this approach, the bank estimates, for each business line/risk type cell, the probability distributions of the severity (single event impact) and of the one year event frequency using its internal data. With these two distributions, the bank then computes the probability distribution of the aggregate operational loss. The total required capital is the sum of the *Value-at-Risk* of each business line and event type combination.

The aim of this paper is threefold:

- First, we develop a sound and rather exhaustive methodological framework in order to support **LDA**, which is seen — in the document [1] — as the ultimate goal of a regulatory incentive-based process. Although **LDA** is not yet allowed, probably because only few banks are able to implement it, there are no doubts that **LDA** will be a matter of great concern for all financial institutions.
- Secondly, we focus on statistical issues related to the available data. Indeed, from a methodological point of view, **LDA** may appear as less complicated to build than internal models for credit risk (or market risk). In line with credit risk models, one has to compute a mathematical mix of two probability distributions: one for the frequency of events and one for the severity. However, contrary to credit risk methodology, the two underlying distributions do not need to be as sophisticated. Nevertheless, the quality and the quantity of data are of greater concern when dealing with operational risk, as the available data could be rare and/or of poor quality. Furthermore, an operational event (i.e. a default in the credit risk vocabulary) is often endogenous as it is related to the internal processes of the institution. As a result, the use of internal data (optimally combined with external ones) is an essential requirement if one wants to obtain a sound, incentive-oriented, internal model. Practically, it is thus necessary to have a careful statistical treatment of all shortcomings around data, such as biased, poor-quality and non-representative data.
- Thirdly, we try to fill the gap between **LDA** and **IMA**. As far as we understand, **IMA** is an attempt to mimic **LDA** through a simplified, easy-to-implement way. While **LDA** requires the use of disaggregated data (on a single event basis), the implementation of **IMA** would be based on aggregated data like the number of events, the total loss by event type, an exposition index, etc. To ensure a good approximation of **LDA** by **IMA**, a gamma factor and a risk profile index — which remains to be precisely defined — would be added on. The third part of the paper is dedicated to the issue of designing an optimal **IMA** in order to be as close as possible to **LDA** and simultaneously to satisfy the constraints required by an easy implementation.

**Remark 1** *In the following, we use the terminology ‘event type’ which seems to be more appropriate. Nevertheless, we could replace ‘event type’ by ‘loss type’ and that will change nothing (except for the categorization).*

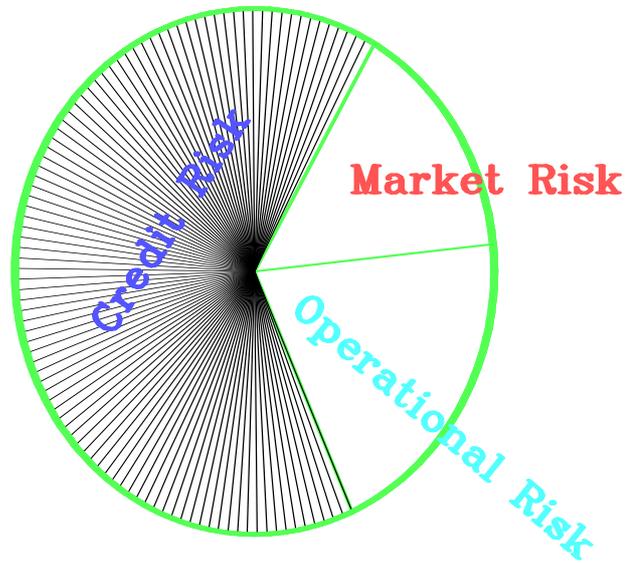


Figure 1: The different categories of risk

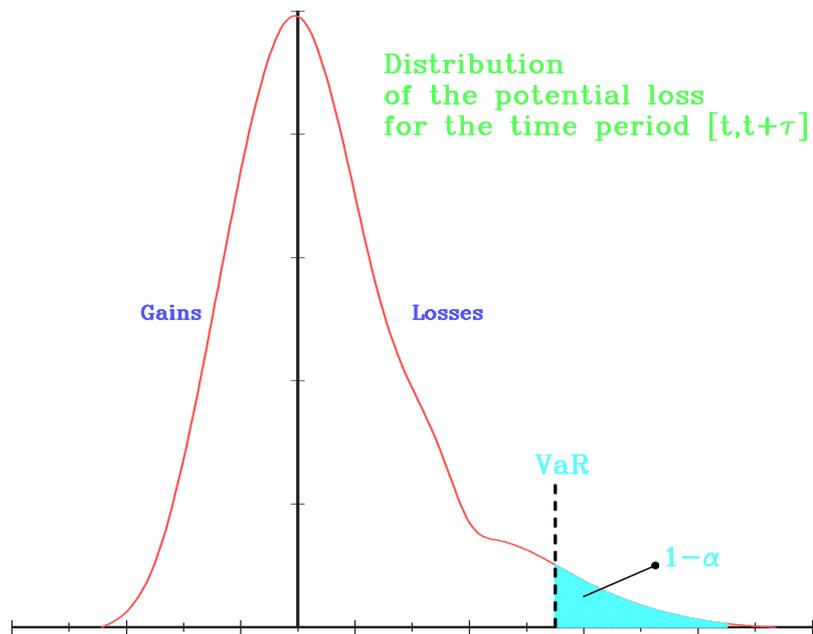


Figure 2: Loss distribution and *Value-at-Risk*

## 2 Loss Distribution Approach

**LDA** is a statistical approach which is very popular in actuarial sciences for computing aggregate loss distributions<sup>1</sup>. In this section, we define the underlying mathematical model<sup>2</sup>, we give some algorithms to compute the distribution and show how to calculate the capital charge based on a *Value-at-Risk* measurement of risk. Moreover, we consider **LDA** as a framework for bottom-up economic capital allocation.

### 2.1 Computing the loss distribution

**LDA** is presented in appendix 6 of [1]:

*Under the Loss Distribution Approach, the bank estimates, for each business line/risk type cell, the probability distribution functions of the single event impact and the event frequency for the next (one) year using its internal data, and computes the probability distribution function of the cumulative operational loss.*

In order to present the mathematical formulation of **LDA**, we have to define some concepts:

- According to the New Basel Capital Accord, we consider different business lines and event types. We use the indices  $i$  and  $j$  to denote a given business line and a given event type.
- $\zeta(i, j)$  is the random variable which represents the **amount of one loss event** for the business line  $i$  and the event type  $j$ . The **loss severity distribution** of  $\zeta(i, j)$  is denoted by  $\mathbf{F}_{i,j}$ .
- We assume that the number of events between times  $t$  and  $t + \tau$  is random<sup>3</sup>. The corresponding variable  $N(i, j)$  has a probability function  $p_{i,j}$ . The **loss frequency distribution**  $\mathbf{P}_{i,j}$  corresponds to

$$\mathbf{P}_{i,j}(n) = \sum_{k=0}^n p_{i,j}(k) \quad (1)$$

In **LDA**, the loss for the business line  $i$  and the event type  $j$  between times  $t$  and  $t + \tau$  is

$$\vartheta(i, j) = \sum_{n=0}^{N(i,j)} \zeta_n(i, j) \quad (2)$$

Let  $\mathbf{G}_{i,j}$  be the distribution of  $\vartheta(i, j)$ .  $\mathbf{G}_{i,j}$  is then a compound distribution

$$\mathbf{G}_{i,j}(x) = \begin{cases} \sum_{n=1}^{\infty} p_{i,j}(n) \mathbf{F}_{i,j}^{n\star}(x) & x > 0 \\ p_{i,j}(0) & x = 0 \end{cases} \quad (3)$$

where  $\star$  is the *convolution*<sup>4</sup> operator on distribution functions and  $\mathbf{F}^{n\star}$  is the  $n$ -fold convolution of  $\mathbf{F}$  with itself<sup>5</sup>.

<sup>1</sup>In fact, this is a very old tool (see for example the section 2.3 of BÜHLMANN [1970]).

<sup>2</sup>Probabilistic aspects are presented in GRANDELL [1991].

<sup>3</sup> $\tau$  is generally equal to one year.

<sup>4</sup>We use the following definition for convolution:

**Definition 2 (Feller [1971, definition 1, p. 143])** *The convolution of a bounded point function  $\varphi$  with a probability distribution  $\mathbf{F}$  is the function defined by*

$$u(x) = \int_{-\infty}^{+\infty} \varphi(x-y) \mathbf{F}(dy) \quad (4)$$

*It will be denoted by  $u = \mathbf{F} \star \varphi$ . When  $\mathbf{F}$  has a density  $f$  we write alternatively  $u = f \star \varphi$  and we have*

$$u(x) = \int_{-\infty}^{+\infty} \varphi(x-y) f(y) dy \quad (5)$$

<sup>5</sup>We have

$$\begin{aligned} \mathbf{F}^{1\star} &= \mathbf{F} \\ \mathbf{F}^{n\star} &= \mathbf{F}^{(n-1)\star} \star \mathbf{F} \end{aligned} \quad (6)$$

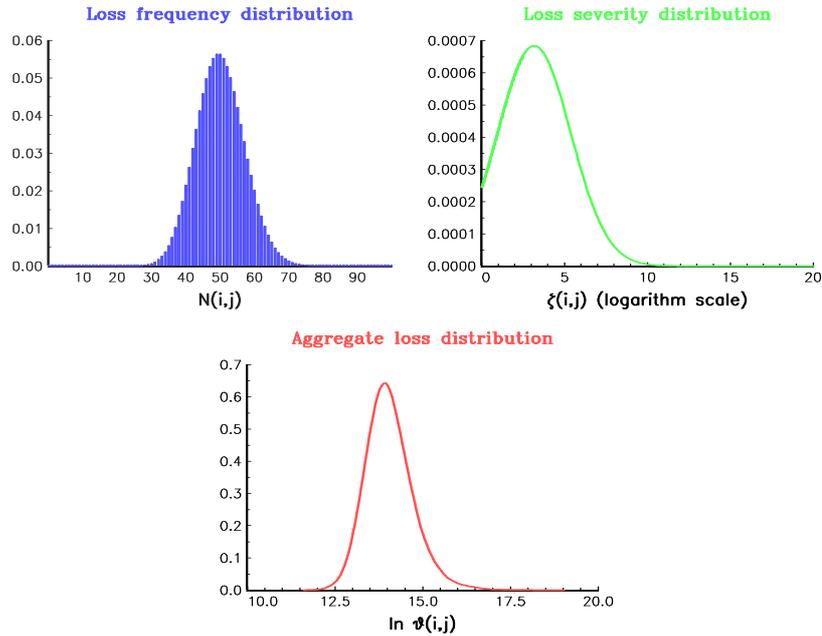


Figure 3: Computing the aggregate loss distribution

**Remark 3** In LDA, the modelling of the aggregate loss distribution is done in two steps. We first consider a loss severity distribution and a loss frequency distribution. Then, we obtain the aggregate loss distribution by compounding them. We have illustrated this method in Figure 3 with  $\zeta(i, j) \sim \mathcal{LN}(8, 2.2)$  and  $N(i, j) \sim \mathcal{P}(50)$ .

**Remark 4** We assume implicitly that the random variables  $\zeta(i, j)$  are independently distributed and independent of the number of events.

In general, there is no analytical expression of the compound distribution<sup>6</sup>. Computing the loss distribution requires then numerical algorithms. The most used are the Monte Carlo method, the Panjer's recursive approach and the inverse of the characteristic function<sup>7</sup>.

- In the Monte Carlo method, the distribution  $\mathbf{G}_{i,j}$  is then 'approximated' by the set  $\mathcal{S}(\vartheta(i, j)) = \{\vartheta_s(i, j), s = 1, \dots, S\}$  of simulated values of the random variable  $\vartheta(i, j)$ . An estimate of  $\mathbf{G}_{i,j}$  is then obtained by the empirical distribution of  $\mathcal{S}(\vartheta(i, j))$  (FISHMAN [1996]) or by the Kernel method (SILVERMAN [1986]).
- PANJER [1981] introduces recursive approaches to compute high-order convolutions. He shows that if

---

<sup>6</sup>The existence of an analytical expression is related to the infinitely divisible property of the distribution  $\mathbf{G}_{i,j}$ . This is for example the case of the negative binomial distribution, which can be written as a compound Logarithmic/Poisson distribution (FELLER [1968]).

<sup>7</sup>HECKMAN and MEYERS [1983] list two other methods:

- In the first method, we 'approximate' the aggregate loss distribution with an assumed analytical distribution. For example, we can use the Normal Power approximation or the transformed Gamma approximation of MONG [1980] (see also VENTER [1982]).
- The second method consists in the Laplace transform inversion method (CSÖRGÓ and TEUGELS [1990]).

Note that other methods exist (for example, WILLMOT and LIN [2000] propose to use Tijms approximations).

there exist constants  $c_1$  and  $c_2$  such that

$$p_{i,j}(n) = \left(c_1 + \frac{c_2}{n}\right) p_{i,j}(n-1) \quad (7)$$

then the following recursion holds

$$g_{i,j}(x) = p_{i,j}(1) f_{i,j}(x) + \int_0^x \left(c_1 + c_2 \frac{y}{x}\right) f_{i,j}(y) g_{i,j}(x-y) dy \quad (8)$$

SUNDT and JEWELL [1981] show that probability distributions that satisfy (7) are the Poisson, binomial, negative binomial and geometric families. For example, if  $N(i, j) \sim \mathcal{P}(\lambda_{i,j})$ , we obtain

$$g_{i,j}(x) = \lambda_{i,j} e^{-\lambda_{i,j}} f_{i,j}(x) + \frac{\lambda_{i,j}}{x} \int_0^x y f_{i,j}(y) g_{i,j}(x-y) dy \quad (9)$$

- HECKMAN and MEYERS [1983] propose to compute aggregate loss distributions using properties of its characteristic function. Let  $X$  be a random variable with distribution  $\mathbf{H}$ . The characteristic function of  $\mathbf{H}$  is then defined by

$$\phi_{\mathbf{H}}(t) = \mathbb{E} [e^{itX}] = \int_0^\infty e^{itx} d\mathbf{H}(x) \quad (10)$$

We remark that the characteristic function of  $M$  independent random variables is the product of their characteristic functions

$$\begin{aligned} \phi_{\mathbf{H}_1 * \mathbf{H}_2 * \dots * \mathbf{H}_M}(t) &= \mathbb{E} [e^{it(X_1 + X_2 + \dots + X_M)}] \\ &= \prod_{m=1}^M \mathbb{E} [e^{itX_m}] \\ &= \prod_{m=1}^M \phi_{\mathbf{H}_m}(t) \end{aligned} \quad (11)$$

It comes that the characteristic function of  $\mathbf{G}_{i,j}$  is given by<sup>8</sup>

$$\phi_{\mathbf{G}_{i,j}}(t) = \sum_{n=0}^{\infty} p_{i,j}(n) [\phi_{\mathbf{F}_{i,j}}(t)]^n \quad (13)$$

We finally deduce the distribution function using the Laplace transformation:

$$\mathbf{G}_{i,j}(x) = \frac{1}{2} - \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{1}{t} e^{-itx} \phi_{\mathbf{G}_{i,j}}(t) dt \quad (14)$$

In the case where  $\mathbf{F}_{i,j}$  is — or can be approximated by — a piecewise linear function or piecewise uniform function, explicit algorithms to compute  $\mathbf{G}_{i,j}(x)$  can be found in HECKMAN and MEYERS [1983] and ROBERTSON [1992].

## 2.2 Computing the *Capital-at-Risk*

With **LDA**, the capital charge (or the *Capital-at-Risk*) is a *Value-at-Risk* measure of risk. We first consider the CaR computation for a given business line and a given event type, and then the CaR computation for the bank as a whole.

---

<sup>8</sup>For example, if  $N(i, j)$  is a Poisson  $\mathcal{P}(\lambda_{i,j})$  distributed random variable, we obtain

$$\phi_{\mathbf{G}_{i,j}}(t) = e^{\lambda_{i,j} (\phi_{\mathbf{F}_{i,j}}(t) - 1)} \quad (12)$$

# Loss Distribution Approach for operational risk

---

## 2.2.1 For one business line and one event type

The *expected loss*  $\text{EL}(i, j)$  and the *unexpected loss*  $\text{UL}(i, j; \alpha)$  at confidence level  $\alpha$  are then defined by

$$\text{EL}(i, j) = \mathbb{E}[\vartheta(i, j)] = \int_0^\infty x \, d\mathbf{G}_{i,j}(x) \quad (15)$$

and

$$\text{UL}(i, j; \alpha) = \mathbf{G}_{i,j}^{-1}(\alpha) - \mathbb{E}[\vartheta(i, j)] = \inf\{x \mid \mathbf{G}_{i,j}(x) \geq \alpha\} - \int_0^\infty x \, d\mathbf{G}_{i,j}(x) \quad (16)$$

The expected loss corresponds to the expected value of the random variable  $\vartheta(i, j)$  whereas the unexpected loss is the quantile for the level  $\alpha$  minus the mean. Let us now define the *Capital-at-Risk* CaR to be the capital charge for operational risk. The Basle Committee on Banking Supervision proposes to define CaR as the unexpected loss

$$\text{CaR}(i, j; \alpha) = \text{UL}(i, j; \alpha) \quad (17)$$

Nevertheless, this definition is not commonly accepted and some institutions compute CaR as the sum of the expected loss and the unexpected loss

$$\begin{aligned} \text{CaR}(i, j; \alpha) &= \text{EL}(i, j) + \text{UL}(i, j; \alpha) \\ &= \mathbf{G}_{i,j}^{-1}(\alpha) \end{aligned} \quad (18)$$

In this case, the *Capital-at-Risk* is a *Value-at-Risk* measure.

To compute  $\text{EL}(i, j)$  and  $\text{UL}(i, j; \alpha)$ , we can use the algorithms defined above. Moreover, we remark that the expected loss can be computed directly with the following formula

$$\begin{aligned} \mathbb{E}[\vartheta(i, j)] &= \mathbb{E}[\mathbb{E}[\vartheta(i, j) \mid N(i, j)]] \\ &= \mathbb{E}[N(i, j)] \times \mathbb{E}[\zeta(i, j)] \end{aligned} \quad (19)$$

The determination of  $\text{UL}(i, j; \alpha)$  **with a high accuracy** is difficult. **The previous numerical algorithms induce some errors and the estimated quantile can be very far from the true quantile especially when the severity loss distribution is heavy-tailed and the confidence level is high**<sup>9</sup>. That's why it is very important to control the accuracy of  $\mathbf{G}_{i,j}^{-1}(\alpha)$ . This can be done by verifying that the estimated first four moments are closed to the theoretical ones. To illustrate this problem, we consider the example of the Log-normal/Poisson compound distribution where  $\zeta \sim \mathcal{LN}(\mu, \sigma)$  and  $N \sim \mathcal{P}(\lambda)$ . We compute the aggregate loss distribution by the Monte Carlo method for different numbers  $S$  of simulations. In Figures 4 and 5, we have reported the density of the point estimators of the mean, the standard deviation and the quantiles at 90% and 99% confidence levels<sup>10</sup>. We remark the influence of the parameter  $\sigma$  on the accuracy of the estimates. However in operational risk, the parameters  $\mu$  and  $\sigma$  can take very large values, which in turn require a great number of simulations to achieve a good accuracy. Let us give a second illustration based on the estimate of the standard deviation of the aggregate loss (denoted by  $\sigma[\vartheta]$ ). Figures 6 and 7 give the ratio  $\hat{\sigma}[\vartheta]/\sigma[\vartheta]$  where  $\hat{\sigma}[\vartheta]$  is the empirical estimate of  $\sigma[\vartheta]$ , computed through the Monte Carlo scheme. We see that the route to convergence (i.e.  $\hat{\sigma}[\vartheta]/\sigma[\vartheta] \rightarrow 1$ ) is strongly influenced by the value of  $\sigma$ . As a conclusion, good accuracy is a crucial point and relies on (depending on which of the three methods is used):

- the number of simulations in the Monte Carlo method,
- the number of grid points to define the aggregate loss in the Panjer algorithm,
- and the number of knots to perform the numerical integration in the characteristic function approach.

---

<sup>9</sup>In the Consultative Document on Operational Risk, the Basle Committee on Banking Supervision suggests to take  $\alpha$  equal to 99%. In economic capital project,  $\alpha$  is related to the rating target of the bank. We give here the values of  $\alpha$  which are generally used:

Rating target	BBB	A	AA	AAA
$\alpha$	99.75%	99.9%	99.95%	99.97%

<sup>10</sup>We use an Epanechnikov kernel with 5000 replications to estimate densities. We estimate the quantile with the nonparametric estimator of ZIELIŃSKI [1999].

## Loss Distribution Approach for operational risk

---

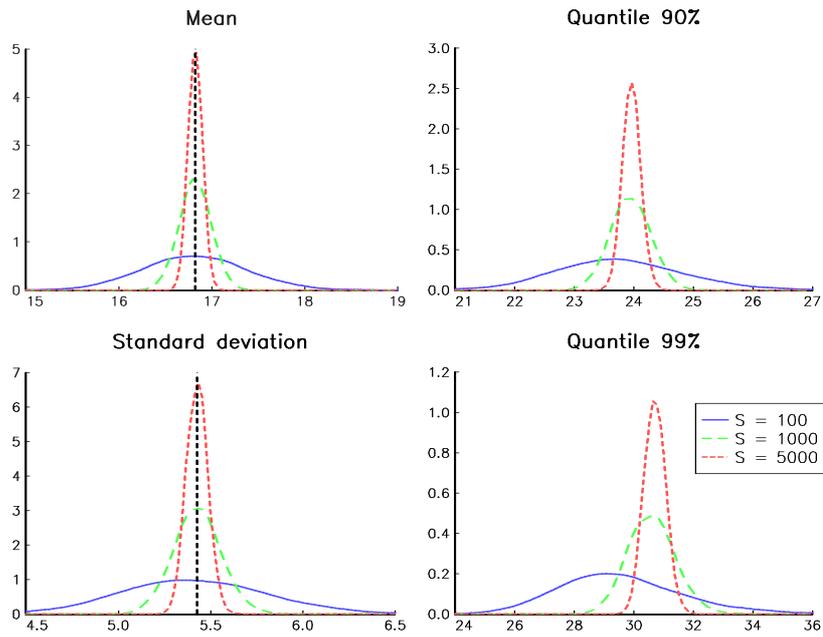


Figure 4: Monte Carlo estimators with  $\lambda = 50$ ,  $\mu = 0.5$  and  $\sigma = 0.2$

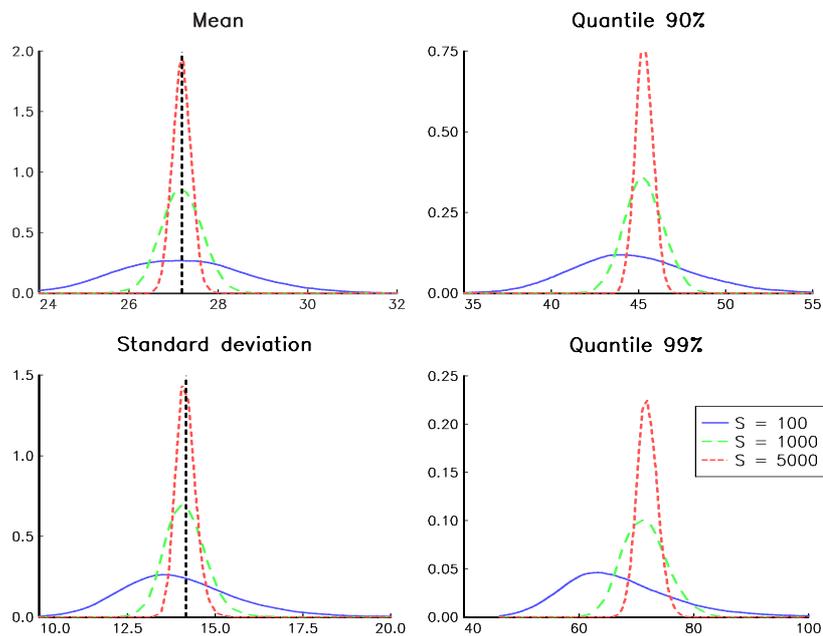


Figure 5: Monte Carlo estimators with  $\lambda = 50$ ,  $\mu = 0.5$  and  $\sigma = 1$

## Loss Distribution Approach for operational risk

---

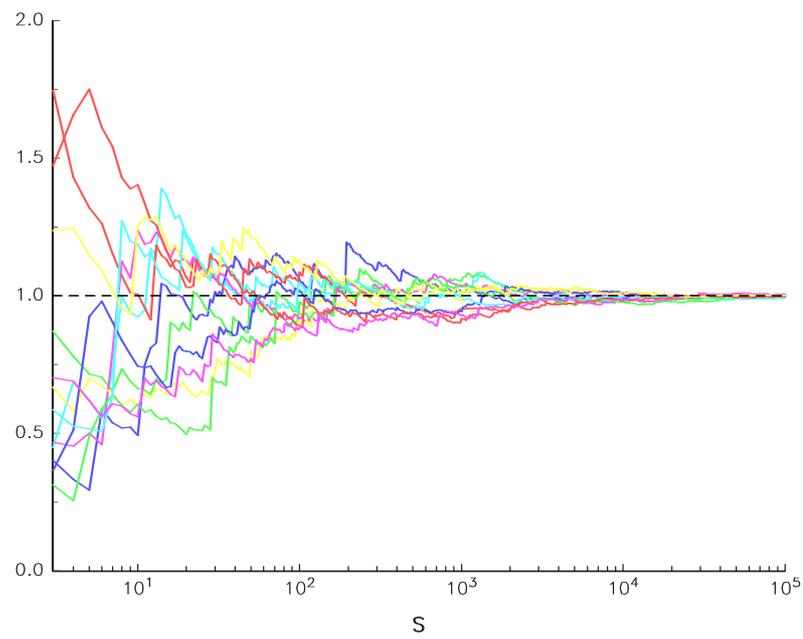


Figure 6: Convergence of the Monte Carlo method with  $\lambda = 10$ ,  $\mu = 5$  and  $\sigma = 1$

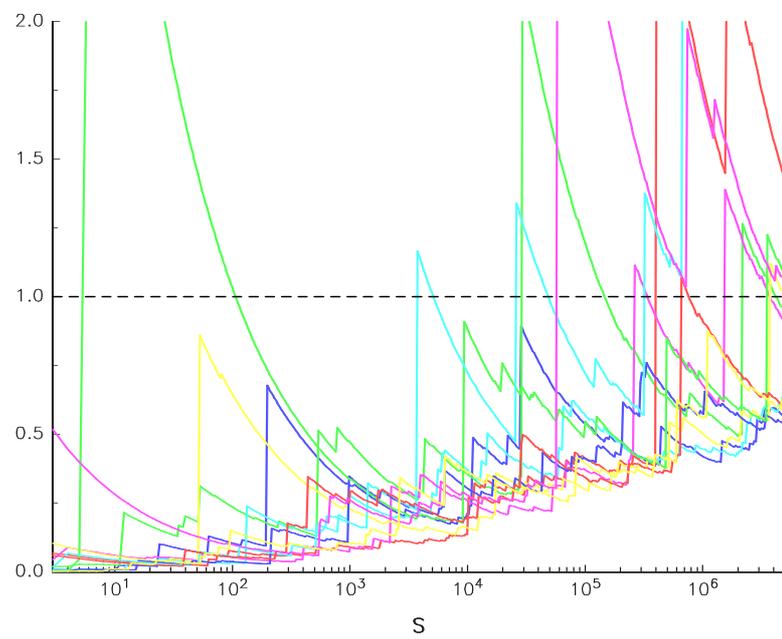


Figure 7: Convergence of the Monte Carlo method with  $\lambda = 10$ ,  $\mu = 5$  and  $\sigma = 3$

### 2.2.2 For the bank as a whole

We consider now the problem of computing the total capital charge for the bank. For example, we may calculate it as the simple summation of the capital charges across each of the business lines and event types:

$$\text{CaR}(\alpha) = \sum_{i=1}^I \sum_{j=1}^J \text{CaR}(i, j; \alpha) \quad (20)$$

This is the method given by the Basel Committee on Banking Supervision in the Internal Measurement Approach. We show in appendix B that it corresponds to the assumption that the different risks are *totally positive dependent*, or roughly speaking, perfectly correlated<sup>11</sup>. A more realistic assumption is to consider that the different losses are independent.

Let  $\vartheta$  be the total loss of the bank:

$$\vartheta = \sum_{i=1}^I \sum_{j=1}^J \vartheta(i, j) \quad (21)$$

If we consider that the losses  $\vartheta(i, j)$  are independent, the distribution  $\mathbf{G}$  is the convolution of the distributions  $\mathbf{G}_{i,j}$ :

$$\mathbf{G}(x) = \star_{i=1}^I \star_{j=1}^J \mathbf{G}_{i,j}(x) \quad (22)$$

As previously defined, the *Capital-at-Risk* of the bank is either:

$$\text{CaR}(\alpha) = \mathbf{G}^{-1}(\alpha) - \mathbb{E}[\vartheta] \quad (23)$$

or

$$\text{CaR}(\alpha) = \mathbf{G}^{-1}(\alpha) \quad (24)$$

Most of the time, the total unexpected loss  $\text{UL}(\alpha)$  is calculated using some approximations. The underlying idea is to define  $\text{UL}(\alpha)$  directly from the individual unexpected loss  $\text{UL}(i, j; \alpha)$  without using the whole distribution  $\mathbf{G}$ . For example, one of the most popular method is the ‘square root rule’ which corresponds to

$$\text{UL}(\alpha) = \sqrt{\sum_{i=1}^I \sum_{j=1}^J \text{UL}^2(i, j; \alpha)} \quad (25)$$

We can show easily that this rule corresponds to a Normal approximation. The main advantage of the formula (25) is its computational tractability<sup>12</sup>. However, it can produce some significant errors. To illustrate this problem, we consider a measure of the diversification effect, called the diversification ratio and defined by

$$\chi(\alpha) = \frac{\text{CaR}^+(\alpha) - \text{CaR}(\alpha)}{\text{CaR}^+(\alpha)} \quad (27)$$

where  $\text{CaR}^+(\alpha)$  is the *Capital-at-Risk* given by the expression (20). Note that we have<sup>13</sup>  $0 < \chi(\alpha) \leq 1$ . We have reported in Table 1 the values taken by  $\chi(\alpha)$  for  $\alpha = 99\%$  and  $\alpha = 99.9\%$  with internal Crédit Lyonnais data. We remark that using Normal approximation (or the square root rule) overestimates the diversification effect and so, underestimates the total *Capital-at-Risk* of the bank.

<sup>11</sup>In statistical language, it corresponds to the case where the dependence function — or the copula — is the upper Fréchet bound.

<sup>12</sup>If we define the *Capital-at-Risk* as the quantile, we have

$$\text{CaR}(\alpha) = \text{EL} + \sqrt{\sum_{i=1}^I \sum_{j=1}^J [\text{CaR}(i, j; \alpha) - \text{EL}(i, j)]^2} \quad (26)$$

<sup>13</sup>In fact,  $\text{CaR}^+(\alpha)$  is not an upper bound of *Capital-at-Risk*. Theoretically, they may exist situations where the total capital charge can be greater than the sum of the individual capital charges (see DURRLEMAN, NIKEGHBALI and RONCALLI [2000]).

## Loss Distribution Approach for operational risk

---

<i>Capital-at-Risk</i> definition	$\mathbf{G}^{-1}(\alpha) - \mathbb{E}[\vartheta]$		$\mathbf{G}^{-1}(\alpha)$	
$\alpha$	99%	99.9%	99%	99.9%
Normal approximation	43.9%	40.9%	30.2%	35.9%
Exact computation	40.0%	37.9%	27.5%	33.2%

Table 1: **LDA** diversification ratio with internal Crédit Lyonnais data

### 2.3 Economic capital allocation

At the present time, the Basel Committee on Banking Supervision does not consider **LDA** to compute regulatory capital. However, **LDA** may be a suitable tool for economic capital allocation:

*It is not envisaged that this approach will be available at the outset of the New Basel Capital Accord, but the Committee does not rule out the use of such an approach in the future. The industry is therefore encouraged to continue its work on developing such approaches. Banks may seek to test such approaches for internal capital allocations purposes (document [1], page 26).*

One of the main justifications of this position is certainly the lack of data available for computing a sound **LDA**. **Even if the bank has an exhaustive database of losses, it can hardly be considered as representative of potential extreme losses.** In this case, **LDA** may largely underestimate the economic capital of the bank. Using external data to supplement internal data may be useful. We discuss this point in paragraph 3.1.

Once we are confident about results, **LDA** is the appropriate tool for bottom-up economic capital allocation. Figure 8 presents a possible scheme to perform it. First, we compute the individual *Capital-at-Risk* based on internal and external data. In general, there are not enough data to consider both business lines and event types. That's why banks may prefer to compute *Capital-at-Risk* per event type in order to obtain more robust results. Then, we compute the consolidated economic capital with or without taking into account diversification effects. At this stage, we have to do the allocation in two steps. In the first step, we consider only event types. Perhaps, the simplest method is to allocate the economic capital using the linear contribution method:

$$EC(j; \alpha) = \delta(j; \alpha) \text{CaR}(\alpha) \quad (28)$$

where

$$\delta(j; \alpha) = \frac{\text{CaR}(j; \alpha)}{\sum_{j=1}^J \text{CaR}(j; \alpha)} \quad \text{or} \quad \delta(j; \alpha) = \frac{\sum_{i=1}^I \text{CaR}(i, j; \alpha)}{\sum_{i=1}^I \sum_{j=1}^J \text{CaR}(i, j; \alpha)} \quad (29)$$

More complicated methods can be considered (DELBAEN and DENAULT [2000], TASCHE [2000]), but they are difficult to implement. Finally, mitigation factors such as insurance coverage must be added on. Secondly, we consider both business lines and event types. If individual *Capital-at-Risk* has been computed by business lines, the economic capital for one business line  $i$  and one specific risk  $j$  is then computed according to:

$$EC(i, j; \alpha) = \delta(i, j; \alpha) EC'(j; \alpha) \quad (30)$$

where

$$\delta(i, j; \alpha) = \frac{\text{CaR}(i, j; \alpha)}{\sum_{i=1}^I \text{CaR}(i, j; \alpha)} \quad (31)$$

and where  $EC'$  denotes the economic capital one mitigation factors have been taken into account. In the other cases, allocation per business line can be done by considering exposure indicators.

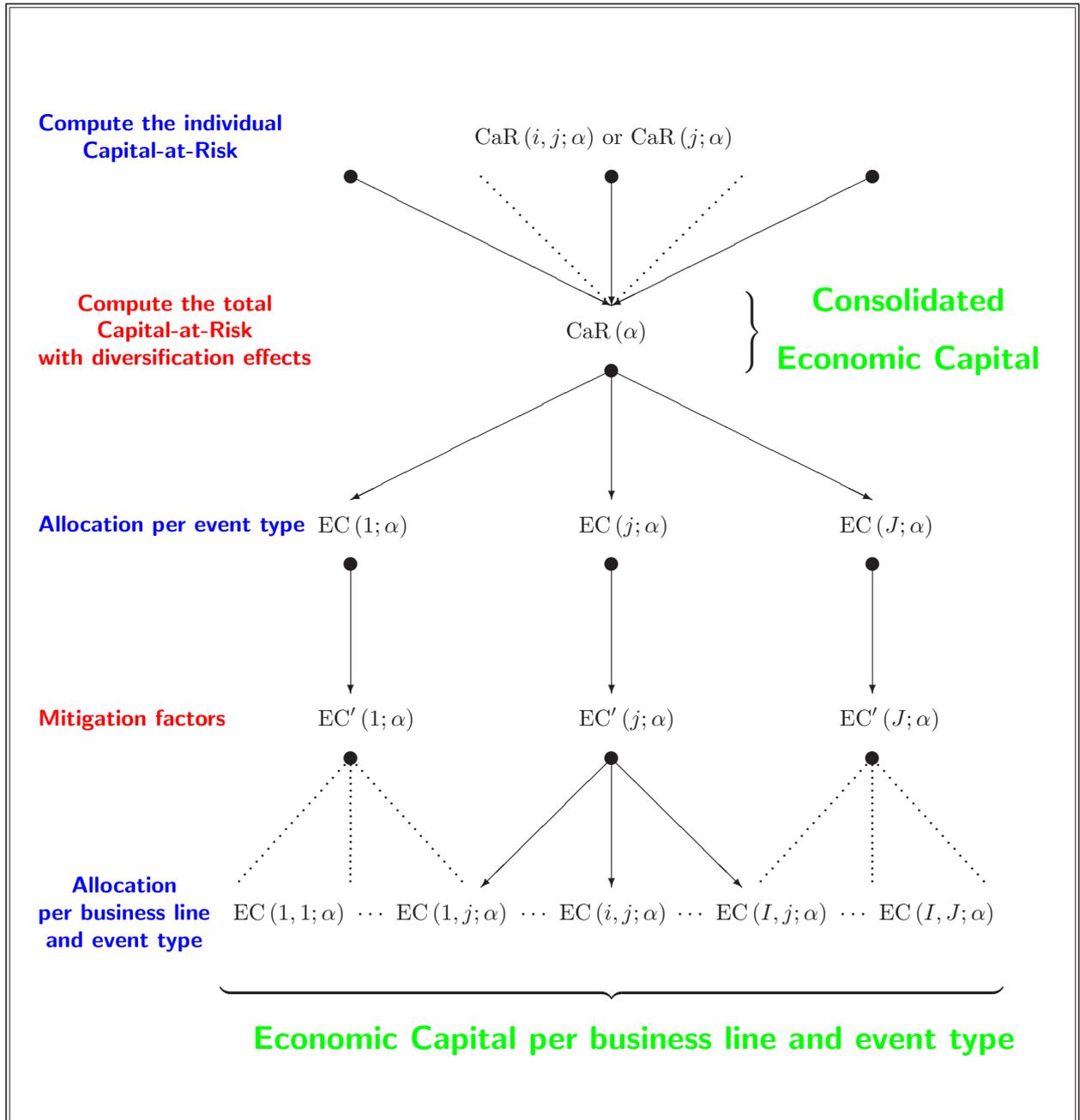


Figure 8: Bottom-up economic capital allocation with LDA

### 3 Some practical issues

In the previous section, we consider the Loss Distribution Approach as a measurement method for operational risk. In this section, we consider some practical issues related to this method. Certainly the most important one relates to the data (available data, exhaustive data and pertinent data). Even with an ideal set of data, LDA requires some expertises.

## 3.1 The data

Data is a crucial issue in operational risk management. It requires an operational risk team and an effective organization to manage the data collection process. As we already mentioned, actual data are often rare and of poor quality leading to the following matters of concern:

- Data are missing for some business lines and/or event types.
- Internal data are biased toward low-severity losses. For obvious reasons, extreme events may be hardly represented in internal databases.
- Only significant losses are reported, meaning that the recorded losses are, by definition, greater than some specific threshold. In statistical terms, this bias is referred as a truncation bias and leads to an over-estimation of the severity. As a matter of fact, one has to find a balance between the cost of recording very low severity data and the truncation bias or accuracy loss resulting from too high thresholds.
- Symmetrically to internal data, external data are biased toward high-severity losses since only they are publicly released. Practically, it means that an appropriate mix between internal and external data must be imposed, in order to enhance statistical efficiency. The current Basel Committee document does not provide any solution to this issue. Our intuition is that it can only be answered by empirical experience. As an example, we have compared the *Capital-at-Risk* for Crédit Lyonnais computed by a benchmark (**B**) and the *Capital-at-Risk* computed with a **LDA** based on our internal data. According to these empirical results, we have to perform a more detailed analysis in order to understand the difference between the two set of results. Some of these differences could be explained by different regulatory or cultural environments and the quality of the internal control. As a result, external data must be used carefully to supplement internal data.

Event type	$\frac{\text{CaR}_{\text{LDA}}(99.9\%)}{\text{CaR}_{\text{B}}(99.9\%)}$
I	1.75
II	0.20
III	0.31
IV	0.59
V	2.20
Total	0.83

Table 2: Comparison of *Capital-at-Risk*

## 3.2 Selection criteria for the severity loss distribution

Estimation methods are presented in the next paragraphs. The problem we address here is the choice of the distribution. Let  $\mathcal{D}$  be the set of probability distribution functions. The question is then the following:

**Which is the ‘best’ distribution in  $\mathcal{D}$  to describe the severity loss distribution?**

First, we have to define the set  $\mathcal{D}$ . We have reported in Table 3 the most used distributions to define the loss severity<sup>14</sup>. Secondly, we have to define a selection criteria. This is generally done using adequacy tests of distributions. For example, we can use non-parametric tests like Anderson-Darling, Cramer-von Mises or Kolmogorov-Smirnov goodness of fit statistics. However, these adequacy tests are not always appropriate for all operational risks, especially for those which are associated to potential extreme losses. Let us illustrate this point. In Figures 9 and 10, we have represented a Quantile-Quantile plot for two estimated distributions  $\mathbf{H}_{(1)}$  and  $\mathbf{H}_{(2)}$ . In both cases, it is clear that  $\mathbf{H}_{(1)}$  fits the data in a better way than  $\mathbf{H}_{(2)}$ . However, there is a difference between the two Figures. Instead, in the second Figure, we observe that  $\mathbf{H}_{(2)}$  fits the upper tail better than  $\mathbf{H}_{(1)}$ .

---

<sup>14</sup>Of course, other distributions could be selected (see [17], [29] or [31] for a more complete list of candidates).

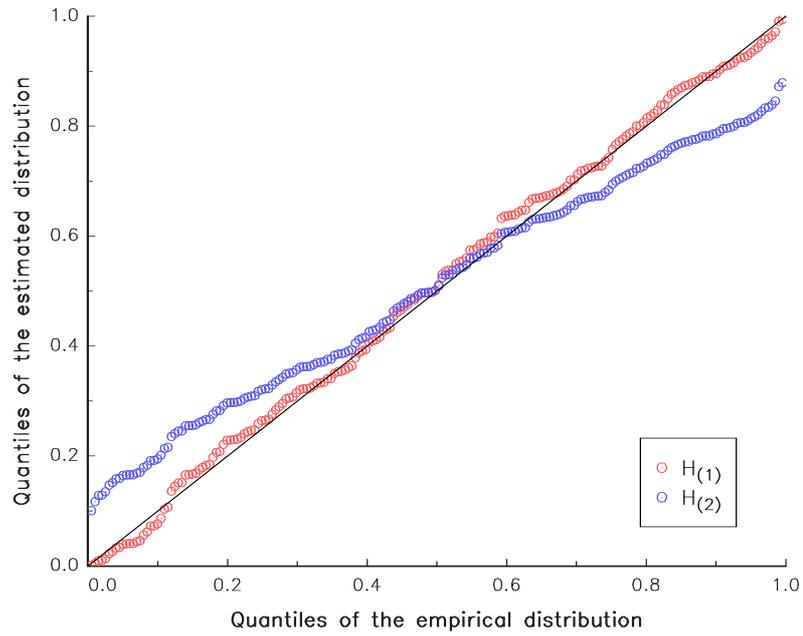


Figure 9: QQ plot without tail problems

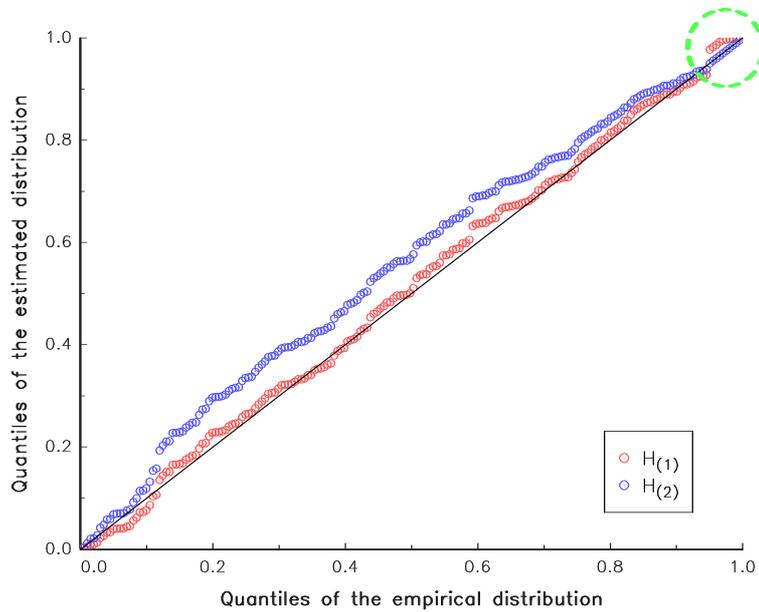


Figure 10: QQ plot with tail problems

## Loss Distribution Approach for operational risk

	Distribution	Expression of the cdf or pdf	Parameters
$\mathcal{LN}$	Log-Normal	$\mathbf{H}(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right)$	$(\mu, \sigma > 0)$
$\mathcal{GEV}$	Generalized extreme value	$\mathbf{H}(x) = \exp\left(-\left(1 + \xi \frac{x - \alpha}{\beta}\right)_+^{-\frac{1}{\xi}}\right)$	$(\alpha, \beta > 0, \xi)$
$\mathcal{GPD}$	Generalized Pareto	$\mathbf{H}(x) = 1 - \left(1 + \xi \frac{x - \alpha}{\beta}\right)_+^{-\frac{1}{\xi}}$	$(\alpha, \beta > 0, \xi)$
$\mathcal{W}$	Weibull	$\mathbf{H}(x) = 1 - \exp\left(-\left(\frac{x - \alpha}{\beta}\right)_+^\xi\right)$	$(\alpha, \beta > 0, \xi)$
$\mathcal{G}$	Gamma	$h(x) = (x - \gamma)_+^{\alpha - 1} [\beta^\alpha \Gamma(\alpha)]^{-1} \exp\left(-\frac{1}{\beta}(x - \gamma)\right)$	$(\alpha > 0, \beta > 0, \gamma > 0)$

Table 3: Some useful distributions

More specifically, we have estimated the parameters of distributions  $\mathcal{LN}$  and  $\mathcal{GEV}$  by maximum likelihood for the event type “Theft and Fraud”. Then we have represented their density functions in Figure 11. The computation of goodness of fit tests does not provide a clear answer to the question of which distribution is the best. However, **the upper tails are different**. Moreover, a diagnostic based on a simplistic graphical approach would be totally misleading as, contrary to what Figure 11 may suggest, the distribution  $\mathcal{GEV}$  has a fatter tail than the  $\mathcal{LN}$  distribution.

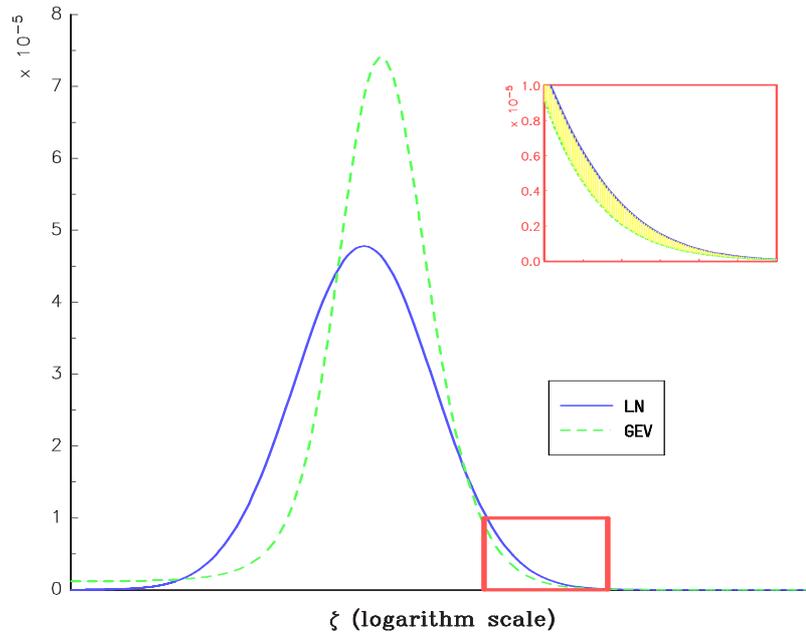


Figure 11: Estimated  $\mathcal{LN}$  and  $\mathcal{GEV}$  distributions for the event type “Theft and Fraud”

In our point of view, goodness of fit statistics are not necessarily the appropriate tool to select a distribution. More pragmatic solutions based on order statistics might be preferred, as they allow to control the upper tail of the distribution. Let  $X_1, \dots, X_n$  be *i.i.d.*  $\mathbf{H}$ -distributed random variables. We define  $X_{m:n}$  as the  $m$ th-order statistic in a sample size  $n$ . We then have

$$X_{1:n} \leq X_{2:n} \leq \dots \leq X_{m:n} \leq \dots \leq X_{n:n} \quad (32)$$

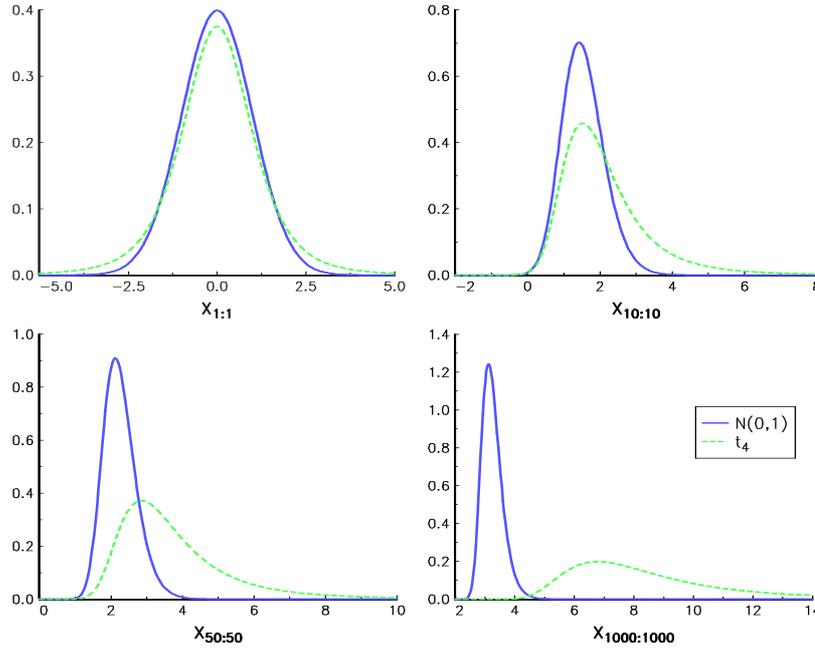


Figure 12: Impact of fat tails on the largest order statistic

Let us denote  $\mathbf{H}_{m:n}$  the distribution of  $X_{m:n}$ . DAVID [1970] shows that the expression of  $\mathbf{H}_{m:n}$  is<sup>15</sup>

$$\mathbf{H}_{m:n}(x) = \sum_{i=m}^n \binom{n}{i} [\mathbf{H}(x)]^i [1 - \mathbf{H}(x)]^{n-i} \quad (34)$$

An important special case is the largest statistic  $X_{n:n}$  which has the distribution  $\mathbf{H}_{n:n}(x) = [\mathbf{H}(x)]^n$  and the density  $h_{n:n}(x) = n [\mathbf{H}(x)]^{n-1} h(x)$ . In order to show the impact of fat tails, we consider the **Normal** and **Student** distributions denoted respectively by  $\mathcal{N}(0,1)$  and  $t_4$ . In Figure 12, we have represented the density function of the corresponding largest order statistic for various values of  $n$ . Even if the two distributions are closed, we remark that the distributions of the maximum  $X_{n:n}$  are very different for larger values of  $n$ . For example, if we consider a sample size of 1000 observations, the probability to have a maximum bigger than, say, five is 0.028% in the case of the **Normal** distribution whereas it is 97.65% in the case of the **Student** distribution!

Considering the previous event type “Theft and Fraud”, Figure 13 gives the distribution of  $\xi_{n:n}$  for  $n$  equal respectively 1, 10, 50 and 100. We remark clearly the impact of the fatter tail of the  $\mathcal{G}\mathcal{E}\mathcal{V}$  distribution. Now, let us fix  $n$  to the number of observations in the internal database. In Figure 14, we have reported the biggest observed losses and the survival distribution of  $\xi_{n:n}$ . Using the  $\mathcal{G}\mathcal{E}\mathcal{V}$  distribution, the probability that the largest order statistic is bigger than the observed largest loss is close to one. For the  $\mathcal{L}\mathcal{N}$  distribution, it is equal to 25%. How to interpret these values? Suppose that you have collected 1500 losses during five years and you will have again 1500 losses in the next five years. There is a probability of one to observe a bigger loss with the  $\mathcal{G}\mathcal{E}\mathcal{V}$  distribution and a probability of 25% with the  $\mathcal{L}\mathcal{N}$  distribution. By using order statistics, we check

<sup>15</sup>We note that the density function  $h_{m:n}(x)$  is

$$h_{m:n}(x) = \frac{n!}{(m-1)!(n-m)!} [\mathbf{H}(x)]^{m-1} [1 - \mathbf{H}(x)]^{n-i} h(x) \quad (33)$$

## Loss Distribution Approach for operational risk

---

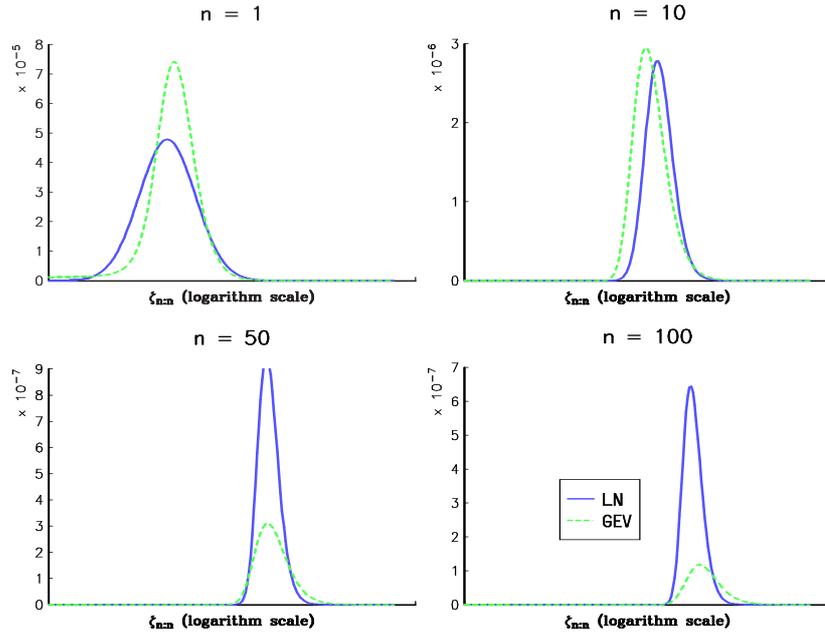


Figure 13: Largest order statistic for the event type “Theft and Fraud”

whether extreme risks are taken into account. For this example, we may conclude that the  $\mathcal{GEV}$  distribution largely overestimates extreme risks.

**Remark 5** In general, the largest order statistic may be used to eliminate distributions which underestimate extreme risks. Let  $x_{m:n}$  be ordered observations. We define  $\varkappa_i$  as follows

$$\varkappa_i = 1 - \mathbf{H}_{n:n}(x_{n-i:n}) \quad (35)$$

$\varkappa_i$  is the probability to observe a maximum greater than the  $i$ th largest ordered observation. Suppose that we use a Normal distribution for the event type “Theft and Fraud”. We obtain the following results

$i$	$\varkappa_i$
1 – 5	$\leq 10^{-5}$
6	12%

It would mean that with a gaussian assumption, the five largest observations should not exist (or had a very small probability to exist).

**Remark 6** Other order statistics may be used, for example the  $i$ th largest order statistic or range of order statistics (see appendix C). They provide a high magnification of the right tail of the estimated distribution. So, when we compute a probability based on these order statistics, we obtain in most cases binary results, either one or either zero. We can immediately verify underestimation or overestimation of extreme risks.

### 3.3 How to deal with aggregated losses?

In this paragraph, we turn to the estimation process. The two best known methods to estimate the parameters of the severity distribution are the method of maximum likelihood and the method of moments.

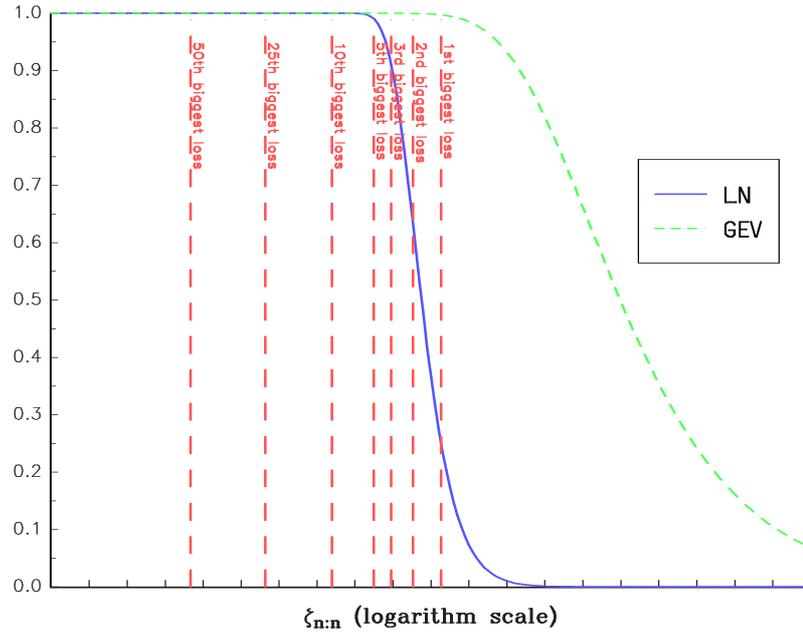


Figure 14: Survival function of the largest order statistic and observed biggest losses

However, both methods are not straightforward because the dataset of losses could contain both events and aggregate of events. When an operational risk dataset is first created, it is not always possible to have a collect of all the past individual events. Most of the times, only sums of losses are available.

A first solution is to drop all the observations of the dataset which do not correspond to an individual loss. Nevertheless, this solution is not satisfactory especially when the dataset contains very few years of historical losses. A second solution consists in taking into account all the observations (individual losses and aggregated losses). We note  $\zeta(i, j)$  the amount of an individual loss and  $\xi(i, j)$  the amount of an aggregated loss. Let us denote the observation by the index  $t$  — do not confuse observations and time. The structure of the dataset is  $(t, n_t(i, j), \xi_t(i, j))$  where  $t$  is the observation,  $n_t(i, j)$  is the number of events corresponding to this observation and  $\xi_t(i, j)$  is the amount of the observed loss. If the observation is an individual loss, then the row of the dataset is  $(t, 1, \zeta_t(i, j))$ . So, individual losses can be considered as a special case of aggregated losses. Mathematically,  $\xi_t(i, j)$  is defined as

$$\xi_t(i, j) = \sum_{n=1}^{n_t(i, j)} \zeta_n(i, j) \quad (36)$$

Let us consider the method of maximum likelihood. It requires the analytical expression of the distribution of  $\xi(i, j)$ , which is not available in most cases. That is why we have to use other estimation methods. For example, we could consider the generalized method of moments. If we know the moments of the severity distribution, we

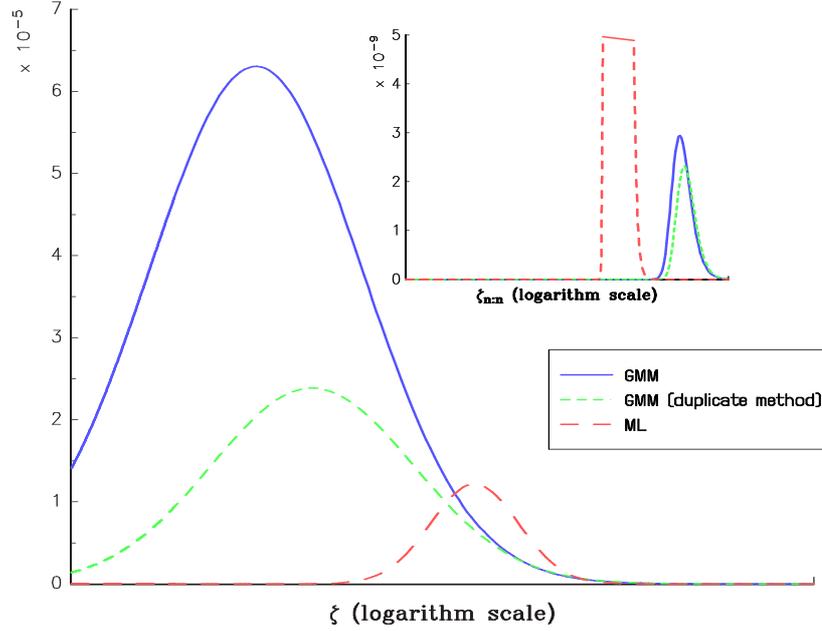


Figure 15: Estimated  $\mathcal{LN}$  distribution for the loss type “Disasters”

could obtain an expression<sup>16</sup> of the moments of  $\xi(i, j)$ :

$$\begin{aligned}
 \mathbb{E}[\xi_t(i, j)] &= n_t(i, j) \mathbb{E}[\zeta(i, j)] \\
 \mathbb{E}[\xi_t(i, j)^2] &= n_t(i, j) \mathbb{E}[\zeta(i, j)^2] + n_t(i, j)(n_t(i, j) - 1) \mathbb{E}^2[\zeta(i, j)] \\
 \mathbb{E}[\xi_t(i, j)^3] &= n_t(i, j) \mathbb{E}[\zeta(i, j)^3] + 3n_t(i, j)(n_t(i, j) - 1) \mathbb{E}[\zeta(i, j)] \mathbb{E}[\zeta(i, j)^2] + \\
 &\quad n_t(i, j)(n_t(i, j) - 1)(n_t(i, j) - 2) \mathbb{E}^3[\zeta(i, j)] \\
 \mathbb{E}[\xi_t(i, j)^4] &= n_t(i, j) \mathbb{E}[\zeta(i, j)^4] + 4n_t(i, j)(n_t(i, j) - 1) \mathbb{E}[\zeta(i, j)] \mathbb{E}[\zeta(i, j)^3] + \\
 &\quad 3n_t(i, j)(n_t(i, j) - 1) \mathbb{E}^2[\zeta(i, j)^2] + \\
 &\quad 6n_t(i, j)(n_t(i, j) - 1)(n_t(i, j) - 2) \mathbb{E}^2[\zeta(i, j)] \mathbb{E}[\zeta(i, j)^2] + \\
 &\quad n_t(i, j)(n_t(i, j) - 1)(n_t(i, j) - 2)(n_t(i, j) - 3) \mathbb{E}^4[\zeta(i, j)]
 \end{aligned} \tag{37}$$

Parameters of the severity distribution are then easily estimated with GMM. Note that more complicated methods of estimation such as the method of simulated moments or indirect inference exist when there are some difficulties to implement GMM (see appendix D).

In Figure 15, we have represented the estimated  $\mathcal{LN}$  distribution for the event type “Disasters” with different methods. The first and second ones are GMM. In the first one, we consider all the observations (individual and aggregated losses) and use the two following moment conditions to estimate the parameters:

$$\begin{cases} h_{t,1}(\mu, \sigma) = \xi_t - n_t e^{\mu + \frac{1}{2}\sigma^2} \\ h_{t,2}(\mu, \sigma) = \left( \xi_t - n_t e^{\mu + \frac{1}{2}\sigma^2} \right)^2 - n_t e^{2\mu + \sigma^2} (e^{\sigma^2} - 1) \end{cases} \tag{38}$$

<sup>16</sup>We use the results of appendix A.

In the second one, we replace rows corresponding to an aggregated loss by  $n_t(i, j)$  rows with loss  $\xi_t(i, j)/n_t(i, j)$ , as it is sometimes proposed. The third method is the maximum likelihood where only individual losses are considered — all the aggregated losses are eliminated. We remark some significant differences<sup>17</sup>.

### 3.4 The frequency distribution

We assume generally that the number of events is a Poisson distributed random variable, i.e.  $N(i, j) \sim \mathcal{P}(\lambda_{i,j})$ . Using the fact that the mean and the variance of a Poisson variate is the frequency parameter  $\lambda$ , it is easy to estimate it. However, we could face some problems because operational risk is at an early age in the industry. The bank has then very few complete years of operational risk events. As a result, it is also necessary to adopt a conservative estimate of  $\lambda_{i,j}$  and use sensitivity analysis to measure the relationship between the capital charge and the frequency parameter. In Figures 16 and 17, we have represented the sensitivity of CaR to  $\lambda$  for the five event types. We remark that  $\Delta(\text{CaR})/\Delta(\lambda)$  is less than one and depends on the confidence level  $\alpha$ . With our data,  $\Delta(\text{CaR})/\Delta(\lambda)$  belongs respectively to  $[\frac{2}{3}, \frac{4}{5}]$  and  $[\frac{1}{2}, \frac{2}{3}]$  for  $\alpha = 99\%$  and  $\alpha = 99.9\%$ .

However, we can not generalized the previous results, because our experience leads us to distinguish two types of operational risks: those with potential high severity losses and those with high frequency/low severity losses<sup>18</sup>. In the first case, we remark that an error in estimating the frequency produces a small impact on the capital charge. In the second case, it could produce very different capital charge. The explanation is the following:

- If the distribution is fat-tailed, the probability of two (or more) extreme loss events is very small (except if the frequency is very high).
- In the other case, the *Capital-at-Risk* is not related to **one** extreme loss event, but to many low losses. The frequency plays also a main role in computing CaR.

## 4 Comparison with the Internal Measurement Approach

At this stage, **LDA** is not considered as an option in the New Basel Capital Accord. Indeed, the regulatory framework for operational risk capital charge proposes three approaches:

### 1. Basic Indicator Approach (**BIA**)

In this case, the operational risk is related to a proxy. The Basel Committee proposes to calculate the capital charge using the gross income

$$\text{CaR} = \alpha \times \text{GI} \tag{39}$$

where GI is the gross income.

### 2. Standardized Approach (**SA**)

In this approach, the bank's activities are divided into standardized business lines. “*Within each business line, the capital charge is calculated by multiplying a bank's broad financial indicator by a ‘beta’ factor*” (§27 of [1]). We have

$$\text{CaR}(i) = \beta(i) \times \text{FI}(i) \tag{40}$$

where  $\text{FI}(i)$  is the financial indicator of business line  $i$ . The total capital charge is then the sum of individual capital charges:

$$\text{CaR} = \sum_{i=1}^I \text{CaR}(i) = \sum_{i=1}^I \beta(i) \times \text{FI}(i) \tag{41}$$

---

<sup>17</sup>However, the number of aggregate losses is very small (less than 4% of the total number of events).

<sup>18</sup>see §11 of [1].

## Loss Distribution Approach for operational risk

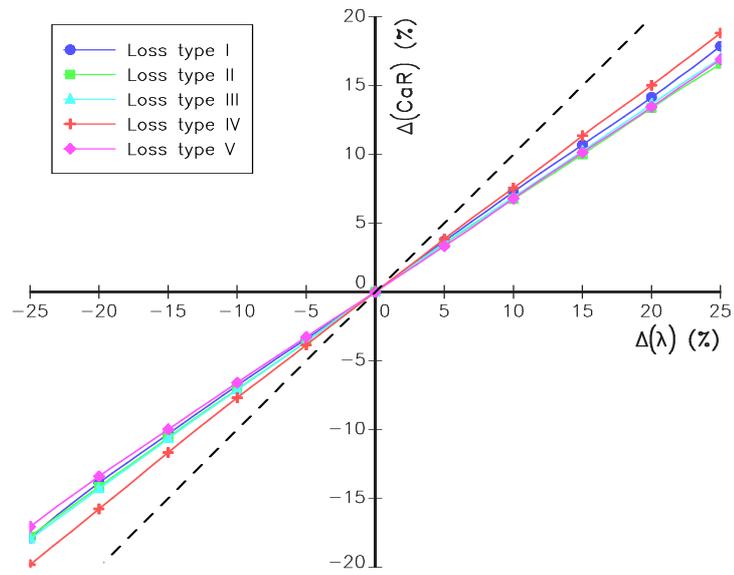


Figure 16: Sensitivity analysis of CaR (99%)

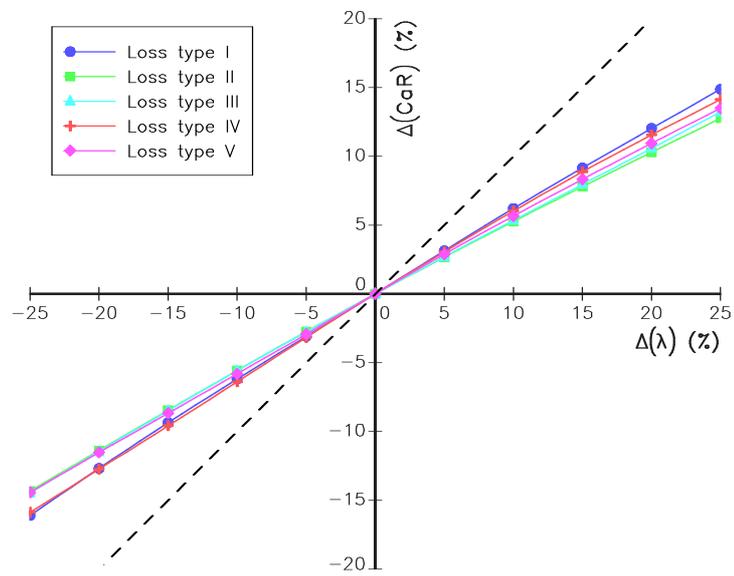


Figure 17: Sensitivity analysis of CaR (99.9%)

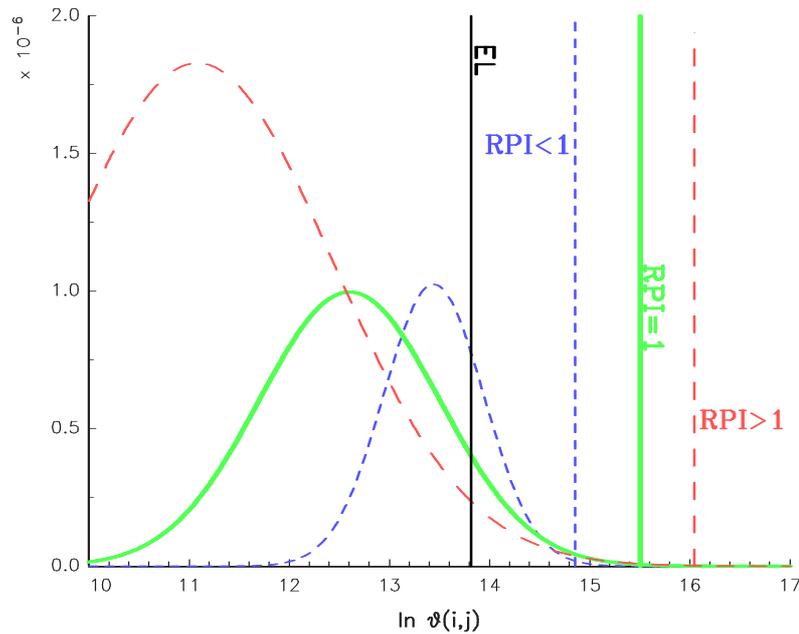


Figure 18: Risk profile index and loss distribution

### 3. Internal Measurement Approach (IMA)

**IMA** is developed “to directly capture a bank’s underlying risk by using the bank’s internal loss data as key inputs for capital calculation” (MORI and HARADA [2001]).

In this last approach, we consider both business lines and event types. For each business line and each event type, the capital charge is computed thanks to the following formula

$$\text{CaR}(i, j) = \text{EL}(i, j) \times \gamma(i, j) \times \text{RPI}(i, j) \quad (42)$$

where  $\text{EL}$  is the expected loss,  $\gamma$  is ‘scaling factor’ and  $\text{RPI}$  is the risk profile index.

- In [1], the Basel Committee on Banking Supervision proposes that the bank estimates the expected loss  $\text{EL}(i, j)$  with the following product

$$\text{EL}(i, j) = \text{EI}(i, j) \times \text{PE}(i, j) \times \text{LGE}(i, j) \quad (43)$$

The different parameters are the exposure indicator  $\text{EI}(i, j)$ , the probability of loss event  $\text{PE}(i, j)$  and the loss given event  $\text{LGE}(i, j)$ .

- The scaling factor  $\gamma(i, j)$  “represents a constant that is used to transform  $\text{EL}$  into risk or a capital charge” (§36 of [1]). This is a regulatory parameter fixed by supervisors.
- The Committee proposes to use a risk profile index  $\text{RPI}(i, j)$  as an adjustment factor to capture the difference of the loss distribution tail of the bank compared to that of the industry wide loss distribution. The idea is to capture the leptokurtic properties of the bank loss distribution and then to transform the

exogeneous scaling factor  $\gamma(i, j)$  into an internal scaling factor  $\gamma^*(i, j)$ :

$$\begin{aligned} \text{CaR}(i, j) &= \text{EL}(i, j) \times \underbrace{\gamma(i, j)}_{\text{Regulatory scaling factor}} \times \text{RPI}(i, j) \\ &= \text{EL}(i, j) \times \underbrace{\gamma^*(i, j)}_{\text{Internal scaling factor}} \end{aligned} \quad (44)$$

It is then obvious that  $\gamma(i, j)$  is a benchmark scaling factor. In Figure 18, we have represented the influence of the loss distribution of the bank on the risk profil index. By definition, the RPI of the industry loss distribution is one. If the bank loss distribution has a fatter tail than the industry loss distribution, then RPI is larger than one. So, two banks which have the same expected loss may have different capital charges because they do not have the same risk profil index.

In what follows, we compare **IMA** and **LDA**. In order to obtain a more tractable **LDA**, we assume that the severity loss distribution is a **Log-Normal**  $\mathcal{LN}(\mu, \sigma)$  distribution whereas the frequency distribution is a **Poisson**  $\mathcal{P}(\lambda)$  distribution. Moreover, we suppose that the *Capital-at-Risk* corresponds to the quantile of the loss distribution, and not to the unexpected loss<sup>19</sup>. In the first paragraph, we consider the scaling problem. The other paragraphs concern the mapping of **LDA** into **IMA** and the definition of RPI.

## 4.1 Scaling the mean or scaling the standard deviation?

The Basel Committee on Banking Supervision has choosen to define the capital charge as a function of the expected loss:

$$\text{CaR} = \gamma^* \times \mathbb{E}[\vartheta] \quad (46)$$

Another solution would be to use the standard deviation instead of the mean:

$$\text{CaR} = \kappa \times \sigma [\vartheta] \quad (47)$$

If we suppose that the Poisson variate is degenerated ( $N = 1$ ), we obtain

$$\gamma^* = \frac{\exp(\sigma\Phi^{-1}(\alpha))}{\exp(\frac{1}{2}\sigma^2)} \quad (48)$$

and

$$\kappa = \frac{\exp(\sigma\Phi^{-1}(\alpha))}{\exp(\sigma)\sqrt{\exp(\sigma^2) - 1}} \quad (49)$$

We remark that the ‘gamma’ and ‘kappa’ factors do not depend on the parameter  $\mu$  (the mean of the logarithm of the losses). Moreover, if we assume that  $\sigma \gg 0$ , it comes that  $\gamma^* = \kappa e^\sigma$ . The relationship between  $\gamma^*$  and  $\kappa$  is non-linear and depends on the value of  $\sigma$  (the standard deviation of the logarithm of the losses).  $\sigma$  is a volatility measure and will influence the kurtosis of the loss distribution. In Figures 19 and 20, we remark that the relationship between  $\gamma^*$  and  $\sigma$  is more obvious than the one between  $\kappa$  and  $\sigma$ . Moreover, for high confidence level, a larger value of  $\sigma$  (or a larger value of kurtosis) implies a larger value of  $\gamma^*$ . This property is not verified for the ‘kappa’ factor.

---

<sup>19</sup>In this case, we have

$$\begin{aligned} \text{CaR}(i, j) &= \text{EL}(i, j) \times [\gamma(i, j) \times \text{RPI}(i, j) - 1] \\ &\simeq \text{EL}(i, j) \times [\gamma(i, j) - 1] \times \text{RPI}(i, j) \end{aligned} \quad (45)$$

## Loss Distribution Approach for operational risk

---

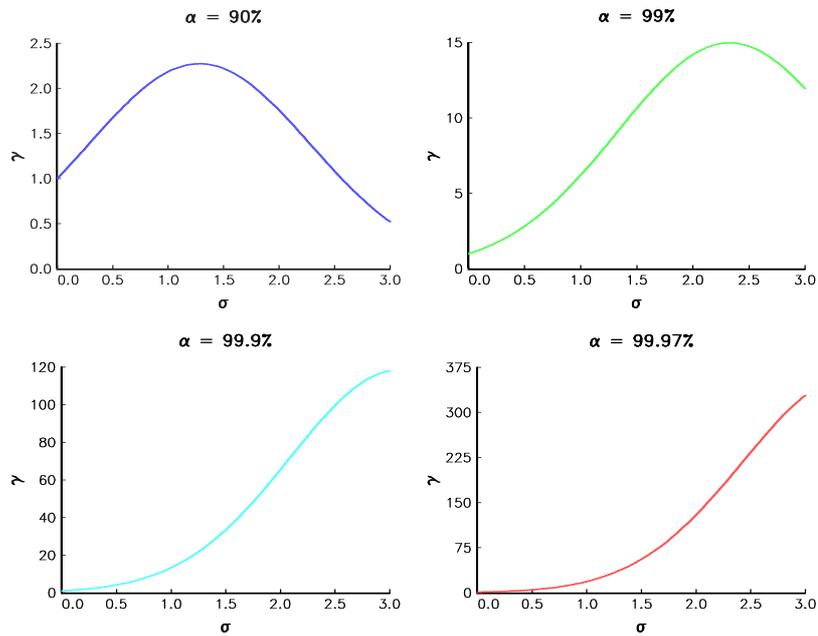


Figure 19: Relationship between the parameter  $\sigma$  and the scaling factor  $\gamma^*$

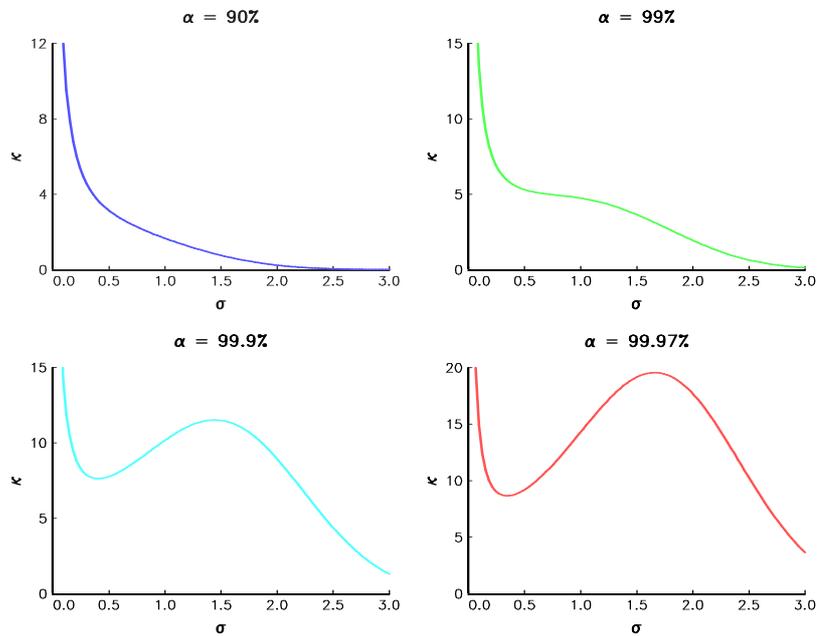


Figure 20: Relationship between the parameter  $\sigma$  and the scaling factor  $\kappa$

### 4.2 Mapping LDA into IMA

In order to write **LDA** as an **IMA** model, we try to find out the conditions under which both methods coincide, i.e.  $EL_{LDA} = EL_{IMA}$  and  $CaR_{LDA} = CaR_{IMA}$ . It comes that the scaling factor has the following expression

$$\begin{aligned}\gamma^* &= \frac{CaR_{IMA}}{EL_{IMA}} \\ &= \frac{CaR_{LDA}}{EL_{LDA}} \\ &= \frac{\mathbf{G}^{-1}(\alpha)}{\lambda \exp\left(\mu + \frac{1}{2}\sigma^2\right)}\end{aligned}\quad (50)$$

where  $\mathbf{G}^{-1}(\alpha)$  is the quantile of  $\vartheta$  for the level  $\alpha$ . Simulations show that  $\mu$  has no influence<sup>20</sup> (or very little) on  $\gamma^*$ . It comes that  $\gamma^*$  depends only on  $\lambda$  and  $\sigma$ :

$$\gamma^* = \gamma^*(\lambda, \sigma; \alpha) \quad (52)$$

We have represented this function in Figures 21 and 22.

It could be useful to have closed-form formula of  $\gamma^*$ . That is why we consider some approximations of the loss distribution. Using a gaussian approximation, it comes that

$$\vartheta \sim \mathcal{N}\left(\lambda \exp\left(m + \frac{1}{2}\sigma^2\right), \sqrt{\lambda} \exp\left(m + \frac{1}{2}\sigma^2\right) \exp\left(\frac{1}{2}\sigma^2\right)\right) \quad (53)$$

and

$$\gamma^* = 1 + \Phi^{-1}(\alpha) \frac{\exp\left(\frac{1}{2}\sigma^2\right)}{\sqrt{\lambda}} \quad (54)$$

In the case of the log-normal approximation, we obtain

$$\vartheta \sim \mathcal{LN}\left(m + \frac{1}{2}\sigma^2 + \frac{3}{2}\ln\lambda - \frac{1}{2}\ln(\lambda + \exp(\sigma^2)), \sqrt{\ln(\lambda^{-1}\exp(\sigma^2) + 1)}\right) \quad (55)$$

and

$$\gamma^* = \sqrt{\lambda} \exp\left(-\frac{1}{2}\ln(\lambda + \exp(\sigma^2)) + \sqrt{\ln(\lambda^{-1}\exp(\sigma^2) + 1)}\Phi^{-1}(\alpha)\right) \quad (56)$$

In the case of the gaussian approximation, we have these two following properties:

$$\begin{aligned}\forall \lambda, \quad \sigma_2 > \sigma_1 &\implies \gamma_2^* > \gamma_1^* \\ \forall \sigma, \quad \lambda_2 > \lambda_1 &\implies \gamma_2^* < \gamma_1^*\end{aligned}\quad (57)$$

In the case of the log-normal approximation, these properties are **almost** always satisfied (for ‘realistic’ values of  $\lambda$  and  $\sigma$ ). Figures 23 and 24 illustrate them.

To give an idea about the pertinence of these approximations, we compare the values given by these approximations ( $\mathcal{N}$  and  $\mathcal{LN}$ ) with the true values ( $\mathcal{CP}$ ). Results with internal Crédit Lyonnais data are reported in Tables 4 and 5. In general, we remark that these approximations underestimate the scaling factor. Nevertheless, for  $\alpha$  equal to 99%, the differences are not very big for some event types.

---

<sup>20</sup>We have

$$\mathbf{G}^{-1}(\alpha) \simeq e^\mu f(\lambda, \sigma; \alpha) \quad (51)$$

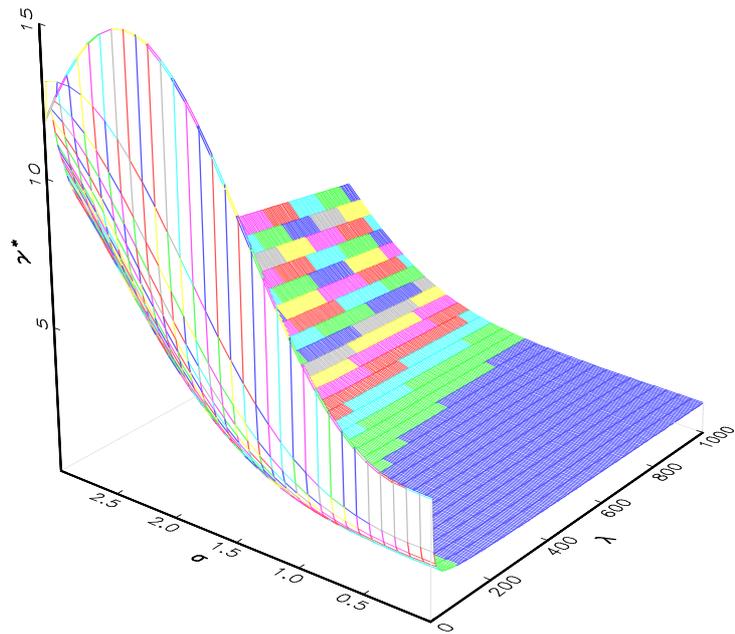


Figure 21: Relationship between  $\lambda$ ,  $\sigma$  and  $\gamma^*$  ( $\alpha = 99\%$ )

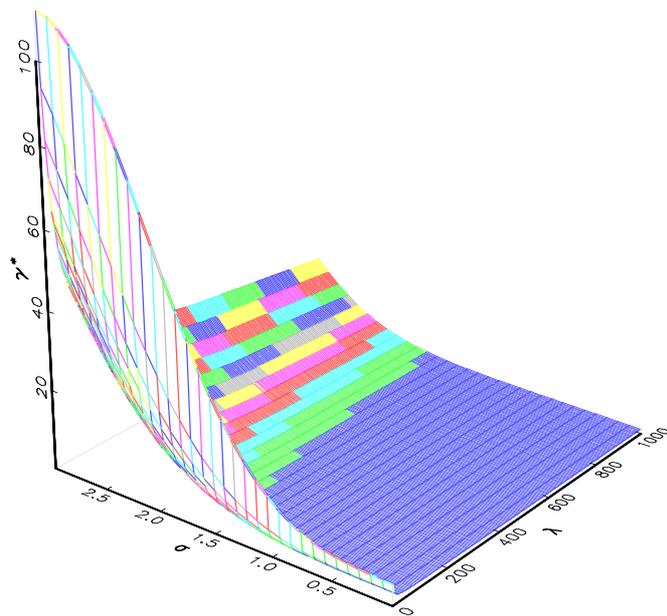


Figure 22: Relationship between  $\lambda$ ,  $\sigma$  and  $\gamma^*$  ( $\alpha = 99.9\%$ )

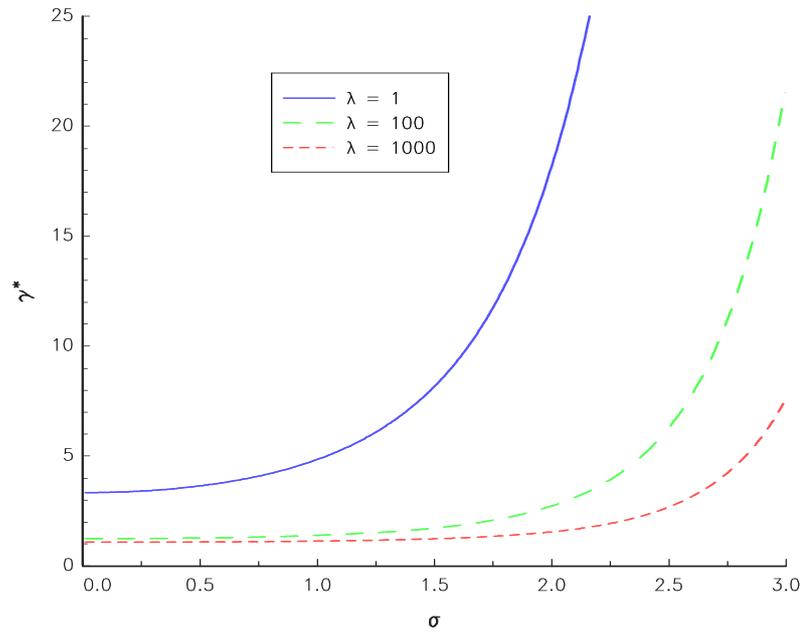


Figure 23: Gaussian approximation of the scaling factor  $\gamma^*$

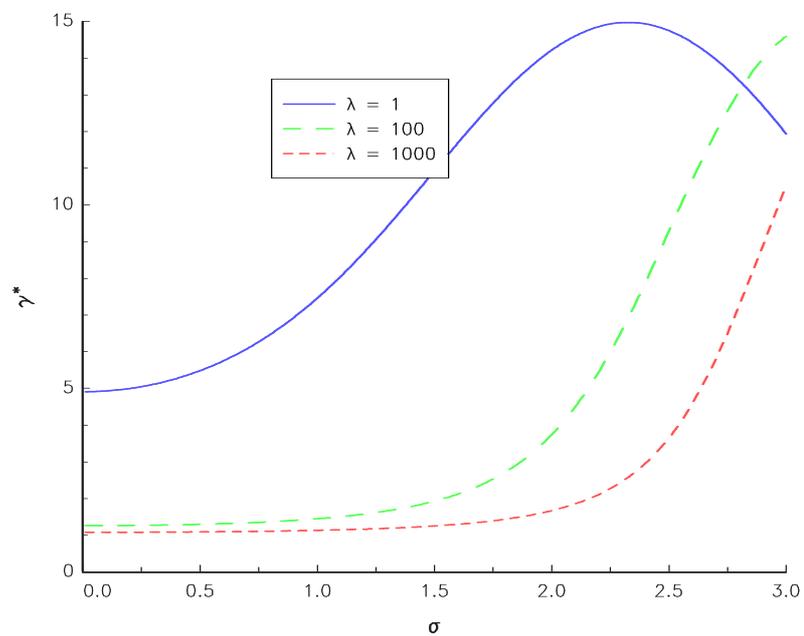


Figure 24: Log-normal approximation of the scaling factor  $\gamma^*$

Event type	$\mathcal{CP}$	$\mathcal{N}$	$\mathcal{LN}$
I	5.5	6.3	9.3
II	2.5	2.0	2.5
III	2.3	1.9	2.2
IV	2.0	1.7	2.0
V	3.5	2.9	4.1

Table 4: Implied ‘gamma’ factors ( $\alpha = 99\%$ )

Event type	$\mathcal{CP}$	$\mathcal{N}$	$\mathcal{LN}$
I	18.4	8.1	26.1
II	4.5	2.4	3.4
III	3.9	2.2	2.9
IV	3.8	2.0	2.5
V	8.5	3.5	7.1

Table 5: Implied ‘gamma’ factors ( $\alpha = 99.9\%$ )

### 4.3 How to define the risk profile index?

**IMA** is supposed to capture the essence of **LDA** in a simplified way. So **IMA** must be designed as a proxy for **LDA** and then should satisfy some constraints. First, it should provide closed-form mathematical expressions, contrary to **LDA** where the unexpected loss is computable only through Monte Carlo simulations or numerical integration. Secondly, **IMA** should be allowed to depend only on aggregate information such as the number of events, the total loss amount (by type of risk and business unit) but not on individual losses. Indeed, the development of **IMA** as a proxy for **LDA** can be justified only if one refuses to use individual losses.

We first note that we can write the equation (44) as

$$\text{CaR} = \text{EL} \times \gamma \times \frac{\gamma^*}{\gamma} \quad (58)$$

RPI could then be interpreted as the ratio between the bank scaling factor and the industry scaling factor. We can then calculate it in two different ways. First, we have

$$\text{RPI} = \frac{\gamma^*}{\gamma} = \frac{\gamma^*(\lambda, \sigma; \alpha)}{\gamma} = \phi_\alpha(\lambda, \sigma) \quad (59)$$

In this case, RPI is related to bank parameters  $\lambda$  and  $\sigma$ . Secondly, we could write

$$\text{RPI} = \frac{\gamma^*}{\gamma} = \frac{\gamma^*(\lambda, \sigma; \alpha)}{\gamma^*(\lambda_\bullet, \sigma_\bullet; \alpha)} = \phi_\alpha(\lambda, \lambda_\bullet, \sigma, \sigma_\bullet) \quad (60)$$

where  $\lambda_\bullet$  and  $\sigma_\bullet$  are the industry parameters. The two expressions are equivalent, but explain the different propositions to define the mathematical formula of RPI.

Using the gaussian approximation, we have

$$\text{RPI} = \frac{1 + \Phi^{-1}(\alpha) \frac{\exp(\frac{1}{2}\sigma^2)}{\sqrt{\lambda}}}{1 + \Phi^{-1}(\alpha) \frac{\exp(\frac{1}{2}\sigma_\bullet^2)}{\sqrt{\lambda_\bullet}}} \quad (61)$$

This formula could not be an **IMA** RPI formulation, because the main idea of **IMA** is to provide a ‘risk sensitive’ method to banks without doing statistical estimations. For example, if we consider the previous RPI expression,

## Loss Distribution Approach for operational risk

---

we have to find an approximation based on indicators which are easy to compute. If we assume that  $\sigma$  is high and  $\lambda$  is low, we have

$$\text{RPI} = \sqrt{\frac{\lambda_{\bullet}}{\lambda}} \times \frac{\exp\left(\frac{1}{2}\sigma^2\right)}{\exp\left(\frac{1}{2}\sigma_{\bullet}^2\right)} \quad (62)$$

An **IMA** RPI formulation could then be

$$\text{RPI}_{\text{IMA}} = \text{RPI}_f \times \text{RPI}_s \quad (63)$$

where  $\text{RPI}_f$  and  $\text{RPI}_s$  are respectively a ‘frequency’ profile index and a ‘severity’ profile index. For example, we could take

$$\text{RPI}_f = \sqrt{\frac{\bar{N}_{\bullet}}{\bar{N}}} \quad (64)$$

and

$$\text{RPI}_s = \frac{\omega}{\omega_{\bullet}} \quad (65)$$

where  $\bar{N}$  is the number of losses for the current year (or an average annual number of losses) and  $\omega$  is the ratio between the mean of losses and the median of losses<sup>21</sup>. So, the idea is then to obtain a **tractable and robust formula**.

Let us consider the different formulas which have been proposed. The I.I.F. Taskforce on “technical” I.M.A. issues proposes to define RPI as

$$\text{RPI}_{\text{IIF}} = \sqrt{\frac{\text{PE}_{\bullet}}{\text{PE}}} \times \frac{\text{EI}_{\bullet}}{\text{EI}} \quad (67)$$

We could say that

$$\text{RPI}_f = \sqrt{\frac{\text{PE}_{\bullet}}{\text{PE}}} \quad (68)$$

and

$$\text{RPI}_s = \sqrt{\frac{\text{EI}_{\bullet}}{\text{EI}}} \quad (69)$$

Using our framework, we remark that  $\text{RPI}_f$  could be viewed as a proxy of  $\sqrt{\frac{\lambda_{\bullet}}{\lambda}}$ . The relationship between our ‘severity’ profile index and  $\sqrt{\frac{\text{EI}_{\bullet}}{\text{EI}}}$  is less evident. Another proposition came from the Sakura Bank Ltd, which specifies

$$\text{RPI}_{\text{SBL}} = 1 + c \frac{1}{\sqrt{N}} \frac{\sigma [\text{EI}]}{\mathbb{E} [\text{EI}]} \quad (70)$$

The main difference between  $\text{RPI}_{\text{IIF}}$  and  $\text{RPI}_{\text{SBL}}$  is that  $\text{RPI}_{\text{IIF}}$  uses industry parameters and  $\text{RPI}_{\text{SBL}}$  uses only bank parameters.

---

<sup>21</sup>In a  $\mathcal{LN}(\mu, \sigma)$  distribution, the median  $m$  is  $e^{\mu}$ . We deduce that

$$\frac{\mathbb{E} [\zeta]}{m [\zeta]} = \frac{\exp\left(\mu + \frac{1}{2}\sigma^2\right)}{\exp(\mu)} = \exp\left(\frac{1}{2}\sigma^2\right) \quad (66)$$

## Loss Distribution Approach for operational risk

---

It is very difficult to say what is the best formulation of RPI without doing the calibration. In what follows, we propose then another RPI formulation. We remind that we have

$$\frac{\text{CaR}_{\text{LDA}}}{\text{EL}_{\text{LDA}}} = \gamma^*(\lambda, \sigma; \alpha) \quad (71)$$

Regarding the latter, one can proxy  $\lambda$  with the average number of events  $\bar{N}$ , which is a natural estimator. Consequently, the questions concerns mainly the  $\sigma$  parameter. In an ideal world, this parameter should be estimated by solving a ML or GMM program but both methods would require to rely on individual losses. However, one can define the two following aggregate statistics  $\bar{\zeta}$  and  $\zeta_{\max}$  which are respectively the average loss and the largest loss recorded in the database. We shall now suppose that theses aggregate statistics are easy to recover. Thus they will serve as the basic instruments around which the proxy for the CaR / EL ratio is developed. Let  $Z$  be a standardized gaussian random variable. We have

$$\zeta := \varphi(Z) = \exp(\mu + \sigma Z) \quad (72)$$

Using the fact that  $\varphi$  is a non-decreasing function, we have

$$\zeta_{\max} = e^\mu e^{\sigma Z_{\max}} \quad (73)$$

and

$$\bar{\zeta} = e^\mu \left[ \frac{1}{M} \sum_{m=1}^M e^{\sigma Z_m} \right] \quad (74)$$

As a result, it comes that there exists a function  $\psi$  such that

$$\sigma = \psi \left( \frac{\zeta_{\max}}{\bar{\zeta}} \right) \quad (75)$$

and finally

$$\frac{\text{CaR}_{\text{LDA}}}{\text{EL}_{\text{LDA}}} = \Gamma^* \left( \bar{N}, \frac{\zeta_{\max}}{\bar{\zeta}}; \alpha \right) \quad (76)$$

So a solution would be to identify the  $\Gamma^*$  function. For example, we could perform a (log-)linear regression of the following type

$$(\ln) \frac{\text{CaR}}{\text{EL}} = c_1 \cdot \bar{N} + c_2 \cdot \frac{\zeta_{\max}}{\bar{\zeta}} + \varepsilon \quad (77)$$

The ratio  $\frac{\zeta_{\max}}{\bar{\zeta}}$  has a natural interpretation. It is always above 1: more specifically, this ratio is close to 1 when  $\sigma$  is near 0 and conversely. So this ratio is a natural measure of the volatility. Furthermore, we expect coefficient  $c_2$  to be positive in the previous regression equation. At this stage, we do not have explored the  $\Gamma^*$  function any further. This is left for our further research.

## 5 Conclusion

In this paper, we have extensively developed the Loss Distribution Approach for computing the capital charge for operational risk. Moreover, we have given some improvements to control the robustness of results. However, even though all problems are not solved yet, **LDA** is (and will be) a powerful method for banks which “*strongly favour risk sensitive/quantitative methods in this area*”.

### References

- [1] Basel Committee on Banking Supervision, Operational Risk — Consultative Document, Supporting document to the New Basel Capital Accord, January 2001
- [2] Derivation of the IMA formula — RPI calculation, The Sakura Bank, Limited, February 2001
- [3] I.I.F. Taskforce on “technical” I.M.A. issues, March 2001
- [4] BALAKRISHNAN, N. and A. CLIFFORD COHEN [1991], Order Statistics and Inference, Academic Press, San Diego
- [5] BEARD, R.E., T. PENTILKÄINEN and E. PESONEN [1984], Risk Theory — The Stochastic Basis of Insurance, *third edition*, *Monographs on Statistics and Applied Probability*, Chapman and Hall, New York (first edition 1969)
- [6] BOUYÉ, E., V. DURRLEMAN, A. NIKEGHBALI, G. RIBOULET and T. RONCALLI [2000], Copulas for finance — a reading guide and some applications, Groupe de Recherche Opérationnelle, Crédit Lyonnais, *Working Paper* (available from <http://gro.creditlyonnais.fr>)
- [7] BOUYÉ, E., V. DURRLEMAN, A. NIKEGHBALI, G. RIBOULET and T. RONCALLI [2001], Copulas: an open field for risk management, Groupe de Recherche Opérationnelle, Crédit Lyonnais, *Working Paper* (available from <http://gro.creditlyonnais.fr>)
- [8] BÜHLMANN, H. [1970], Mathematical Methods in Risk Theory, *Grundlehren der Mathematischen Wissenschaften in Einzeldarstellungen*, **Band 172**, Springer-Verlag, Heidelberg
- [9] CESKE, R. and J.V. HERNÁNDEZ [1999], Where theory meets practice, *Risk Magazine (Operational Risk Report)*, **12**, November, 17-20
- [10] CESKE, R., J.V. HERNÁNDEZ and L.M. SÁNCHEZ [2000], Quantifying event risk: the next convergence, *The Journal of Risk Finance*, **1(3)**, 9-23
- [11] CSÖRGÓ, S. and J.F. TEUGELS [1990], Empirical Laplace transform and approximation of compound distributions, *Journal of Applied Probability*, **27**, 88-101
- [12] DAVID, H.A. [1970], Order Statistics, *Wiley Series in Probability and Mathematical Statistics*, John Wiley & Sons, New York
- [13] DAVIDSON, R. and J.G. MACKINNON [1993], Estimation and Inference in Econometrics, Oxford University Press, Oxford
- [14] DELBAEN, F. and M. DENAULT [2000], Coherent allocation of risk capital, RiskLab, *Working Paper*
- [15] DUFFIE, D. and K.J. SINGLETON [1993], Simulated moments estimation of Markov models of asset prices, *Econometrica*, **61**, 929–952
- [16] DURRLEMAN, V., A. NIKEGHBALI, and T. RONCALLI [2000], How to get bounds for distribution convolutions? A simulation study and an application to risk management, Groupe de Recherche Opérationnelle, Crédit Lyonnais, *Working Paper* (available from <http://gro.creditlyonnais.fr>)
- [17] EVANS, M., N. HASTINGS and B. PEACOCK [1993], Statistical Distributions, *second edition*, John Wiley & Sons, New York
- [18] FELLER, W. [1968], An introduction to Probability Theory and Its Applications — Volume I, *third edition*, *Wiley Series in Probability and Mathematical Statistics*, John Wiley & Sons, New York (first edition 1950)
- [19] FELLER, W. [1971], An introduction to Probability Theory and Its Applications — Volume II, *second edition*, *Wiley Series in Probability and Mathematical Statistics*, John Wiley & Sons, New York (first edition 1966)

- [20] FISHMAN, G.S. [1996], Monte Carlo: Concepts, Algorithms and Applications, Springer Series in Operations Research, Springer-Verlag, New York
- [21] FRANK, M.J., R.B. NELSEN and B. SCHWEIZER [1987], Best-possible bounds for the distribution of a sum — a problem of Kolmogorov, *Probability Theory and Related Fields*, **74**, 199-211
- [22] GALAMBOS, J. [1982], The role of functional equations in stochastic model building, *Aequationes Mathematicae*, **25**, 21-41
- [23] GALLANT, A.R. and G. TAUCHEN [1996], Which moments to match?, *Econometric Theory*, **12**, 657-681
- [24] GOURIÉROUX, C., A. MONFORT and E. RENAULT [1993], Indirect Inference, *Journal of Applied Econometrics*, **8**, S85-S118
- [25] GRANDPELL, J. [1991], Aspects of Risk Theory, *Springer Series in Statistics*, Springer-Verlag, New York
- [26] HANSEN, L.P. [1982], Large sample properties of generalized method of moments estimators, *Econometrica*, **50**, 1029-1054
- [27] HECKMAN, P.E. and G.G. MEYERS [1983], The calculation of aggregate loss distributions from claim severity and claim count distributions, *Proceedings of the Casualty Actuarial Society*, **LXX**, 22-61
- [28] JOE, H. [1997], Multivariate Models and Dependence Concepts, *Monographs on Statistics and Applied Probability*, **73**, Chapman & Hall, London
- [29] JOHNSON, N.L., S. KOTZ and N. BALAKRISHNAN [1994], Continuous Univariate Distributions — Volume 1, *second edition*, John Wiley & Sons, New York (first edition 1970)
- [30] JOHNSON, N.L., S. KOTZ and N. BALAKRISHNAN [1997], Discrete Multivariate Distributions, *Wiley Series in Probability and Mathematical Statistics*, John Wiley & Sons, New York
- [31] KLUGMAN, S.A. [1986], Loss distributions, in H.H. Panjer (Ed.), *Actuarial Mathematics*, American Mathematical Society, Providence, Rhode Island
- [32] KLUGMAN, S.A., H.H. PANJER and G.E. WILLMOT [1998], Loss Models: From Data to Decisions, *Wiley Series in Probability and Mathematical Statistics*, John Wiley & Sons, New York
- [33] MARCEAU, E., H. COSSETTE, P. GAILLARDETZ and J. RIOUX [1999], Dependence in the individual risk model, Université Laval Québec, *Working Paper*
- [34] MONG [1980], Estimating aggregate loss probability and increased limit factor, Pricing Property and Casualty Insurance Products (May 11-14, Caribe Hilton, San Juan, Puerto Rico), *Casualty Actuarial Society Discussion Paper Program*, 358-393
- [35] MORI, T. and E. HARADA [2001], Internal Measurement Approach to operational risk capital charge, Bank of Japan, *Discussion Paper*
- [36] NELSEN, R.B. [1999], An Introduction to Copulas, *Lectures Notes in Statistics*, **139**, Springer-Verlag, New York
- [37] PANJER, H.H. [1981], Recursive evaluation of compound distributions, *Astin Bulletin*, **12**, 22-26
- [38] PANJER, H.H. (editor) [1986], Actuarial Mathematics, *Proceedings of Symposia in Applied Mathematics*, **35**, American Mathematical Society, Providence, Rhode Island
- [39] ROBERTSON, J.P. [1992], The computation of aggregate loss distributions, *Proceedings of the Casualty Actuarial Society*, **LXXIX**, 57-133
- [40] SILVERMAN, B.W. [1986], Density Estimation for Statistics and Data Analysis, *Monographs on Statistics and Applied Probability*, **26**, Chapman & Hall, London

- [41] SONG, P. X-K. [2000], Multivariate dispersion models generated from Gaussian copula, *Scandinavian Journal of Statistics*, **27-2**, 305-320
- [42] SUNDT, B. and W.S. JEWELL [1981], Further results on recursive evaluation of compound distributions, *Astin Bulletin*, **12**, 27-39
- [43] TASCHE, D. [1999], Risk contributions and performance measurement, Zentrum Mathematik, TU München, *Working Paper*, July
- [44] VENTER, G. [1983], Transformed Beta and Gamma distributions and aggregate losses, *Proceedings of the Casualty Actuarial Society*, **LXX**, 156-193
- [45] WANG, S. [1998], Aggregation of correlated risk portfolios: models and algorithms, *Proceedings of the Casualty Actuarial Society*, **LXXXV**, 848-939
- [46] WANG, Y.H. and S. JI [1993], Derivations of the compound Poisson distribution and process, *Statistics & Probability Letters*, **18**, 1-7
- [47] WILLIAMSON, R.C [1989], Probabilistic Arithmetic, Ph.D. Thesis, University of Queensland
- [48] WILLMOT, G.E. and X.S. LIN [2000], Lundberg Approximations for Compound Distributions With Insurance Applications, *Lecture Notes in Statistics*, **156**, Springer-Verlag, New York
- [49] ZIELIŃSKI, R. [1999], Best equivariant nonparametric estimator of a quantile, *Statistics & Probability Letters*, **45**, 79-84

## A Moments of the aggregate loss distribution

We consider the random variable  $\vartheta = \sum_{n=0}^N \zeta_n$  where  $N$  and  $\zeta_n$  are respectively the frequency and severity random variables. Let  $X$  be a random variable ( $N$ ,  $\zeta$  or  $\vartheta$ ). We note  $\dot{\mu}_r^X$  the  $r^{\text{th}}$  moment about the origin and  $\mu_r^X$  the  $r^{\text{th}}$  (central) moment about the mean<sup>22</sup>. To calculate the moments, we use the following relationships

$$\begin{aligned}
 \left( \sum_{i=1}^n x_i \right)^2 &= \sum_{i=1}^n x_i^2 + \sum_{i \neq j}^n x_i x_j \\
 \left( \sum_{i=1}^n x_i \right)^3 &= \sum_{i=1}^n x_i^3 + 3 \sum_{i \neq j}^n x_i^2 x_j + \sum_{i \neq j \neq k}^n x_i x_j x_k \\
 \left( \sum_{i=1}^n x_i \right)^4 &= \sum_{i=1}^n x_i^4 + 4 \sum_{i \neq j}^n x_i^3 x_j + 3 \sum_{i \neq j}^n x_i^2 x_j^2 + 6 \sum_{i \neq j \neq k}^n x_i^2 x_j x_k + \sum_{i \neq j \neq k \neq l}^n x_i x_j x_k x_l
 \end{aligned} \tag{78}$$

It comes that

$$\begin{aligned}
 \dot{\mu}_1^\vartheta &= \dot{\mu}_1^N \dot{\mu}_1^\zeta \\
 \dot{\mu}_2^\vartheta &= \dot{\mu}_1^N \dot{\mu}_2^\zeta + \left( \dot{\mu}_2^N - \dot{\mu}_1^N \right) \left[ \dot{\mu}_1^\zeta \right]^2 \\
 \dot{\mu}_3^\vartheta &= \dot{\mu}_1^N \dot{\mu}_3^\zeta + 3 \left( \dot{\mu}_2^N - \dot{\mu}_1^N \right) \dot{\mu}_1^\zeta \dot{\mu}_2^\zeta + \left( \dot{\mu}_3^N - 3\dot{\mu}_2^N + 2\dot{\mu}_1^N \right) \left[ \dot{\mu}_1^\zeta \right]^3 \\
 \dot{\mu}_4^\vartheta &= \dot{\mu}_1^N \dot{\mu}_4^\zeta + 4 \left( \dot{\mu}_2^N - \dot{\mu}_1^N \right) \dot{\mu}_1^\zeta \dot{\mu}_3^\zeta + 3 \left( \dot{\mu}_2^N - \dot{\mu}_1^N \right) \left[ \dot{\mu}_2^\zeta \right]^2 + 6 \left( \dot{\mu}_3^N - 3\dot{\mu}_2^N + 2\dot{\mu}_1^N \right) \dot{\mu}_2^\zeta \left[ \dot{\mu}_1^\zeta \right]^2 + \\
 &\quad \left( \dot{\mu}_4^N - 6\dot{\mu}_3^N + 11\dot{\mu}_2^N - 6\dot{\mu}_1^N \right) \left[ \dot{\mu}_1^\zeta \right]^4
 \end{aligned} \tag{79}$$

---

<sup>22</sup>We have  $\dot{\mu}_r^X = \mathbb{E}[X^r]$  and  $\mu_r^X = \mathbb{E}[(X - \mathbb{E}[X])^r]$ .

It is then easy to derive the central moments or other characteristics like skewness or kurtosis (see section 3.3 of BEARD, PENTILKÄINEN and PESONEN [1984]). In the case of the compound Poisson model, we obtain very simple formulas because we have

$$\begin{aligned}\dot{\mu}_1^N &= \mu_1^N = \mu_2^N = \lambda \\ \dot{\mu}_2^N &= \lambda(\lambda + 1) \\ \dot{\mu}_3^N &= \lambda(\lambda^2 + 3\lambda + 1) \\ \dot{\mu}_4^N &= \lambda(\lambda^3 + 6\lambda^2 + 7\lambda + 1)\end{aligned}\tag{80}$$

## B General methodology to compute the total capital charge for the bank

We remind that the total loss distribution  $\mathbf{G}$  for the bank as a whole is computed as the convolution of the individual aggregate distributions  $\mathbf{G}_{i,j}$ :

$$\mathbf{G}(x) = \star_{i=1}^I \star_{j=1}^J \mathbf{G}_{i,j}(x)\tag{81}$$

In this case, we assume the independence between the random variables  $\vartheta(i, j)$ . We present now two methods to introduce dependence between the different operational risks. In the first one, the dependence is introduced *directly* on the random variables  $\vartheta(i, j)$  whereas in the second one, the dependence is introduced on the frequencies  $N(i, j)$ . Even if they are different, the two methods use both copulas<sup>23</sup>. Similar ideas may be found in WANG [1998] and MARCEAU, COSSETTE, GAILLARDEZ and RIOUX [1999].

### B.1 Correlated aggregate loss distributions

Let  $\check{\vartheta}$  be the vec form of the random variables  $\vartheta(i, j)$ . We have

$$\check{\vartheta}_{i+I \times (j-1)} = \vartheta(i, j)\tag{82}$$

We note  $\check{\mathbf{G}}$  the distribution of the random vector  $\check{\vartheta}$ . By definition, the margins of  $\check{\mathbf{G}}$  are the individual aggregate loss distributions  $\mathbf{G}_{i,j}$ . However, the knowledge of the margins is not sufficient to compute the total loss for the whole bank:

$$\vartheta = \sum_{i=1}^I \sum_{j=1}^J \vartheta(i, j)\tag{83}$$

We need the dependence function or copula  $\mathbf{C} \langle \check{\mathbf{G}} \rangle$ . Of course, if  $\mathbf{C} \langle \check{\mathbf{G}} \rangle$  is the product copula  $\mathbf{C}^\perp$ , we obtain the previous formula (81). In other cases, the distribution of  $\vartheta$  may be computed using Monte Carlo simulations. To illustrate the method, we consider the following example. We have two event types with  $\zeta_1 \sim \mathcal{LN}(1, 1)$  and  $\zeta_2 \sim \mathcal{LN}(1.25, 0.5)$ . The frequency distributions are respectively  $\mathcal{P}(10)$  and  $\mathcal{P}(12)$ . In Figure 25, we have represented the corresponding aggregate loss distributions. Moreover, we have reported the distribution of the total loss  $\vartheta = \vartheta_1 + \vartheta_2$  for different Normal copulas. When the parameter  $\rho$  of the Normal copula is equal to 0, it is equivalent to use the product copula  $\mathbf{C}^\perp$ . When  $\rho$  is less than 0, the Normal copula  $\mathbf{C}$  is smaller than  $\mathbf{C}^\perp$  ( $\mathbf{C} \prec \mathbf{C}^\perp$ ). In this case, we have a negative dependence function. Positive dependence functions correspond to the case  $\rho > 0$ . We remark that the total loss distribution depends strongly on the value of the parameter  $\rho$ . For example, with positive dependence functions, the total loss distribution is moved to the right. Figure 26 presents the *Capital-at-Risk* for different confidence levels  $\alpha$  and different values of  $\rho$ .

The Basel Committee on Banking Supervision suggests to compute the *Capital-at-Risk* for the whole bank as the simple sum of the *Capital-at-Risk* for each business line/event type. We will show that this method

---

<sup>23</sup>We do not provide a mathematical treatment of this subject and we refer interested readers to JOE [1997] or NELSEN [1999] (see also [6] and [7] for financial applications).

## Loss Distribution Approach for operational risk

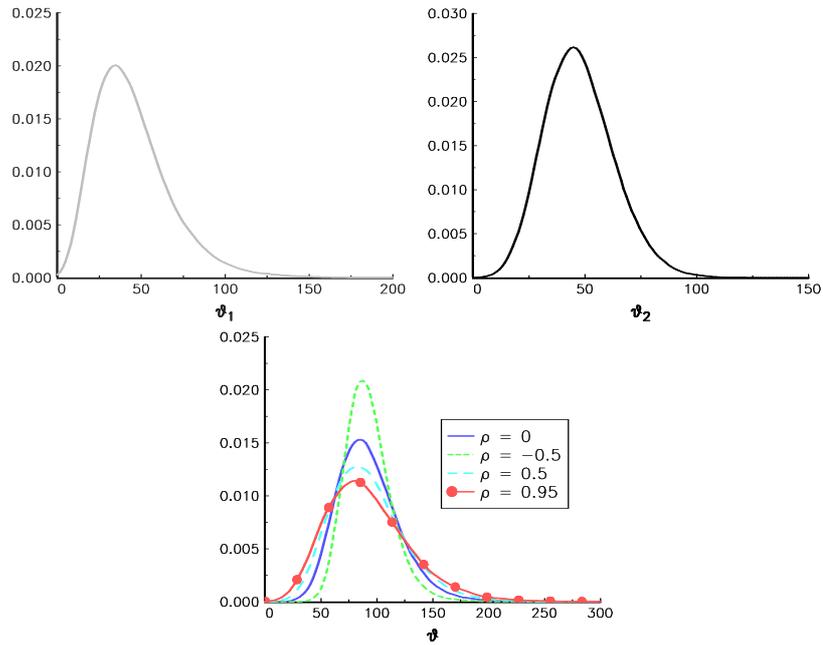


Figure 25: Impact of the dependence function on the total loss distribution

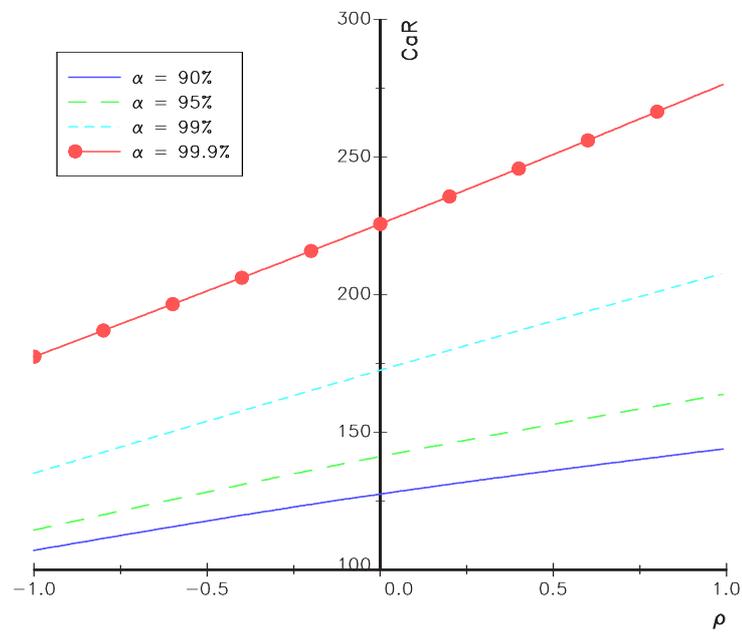


Figure 26: Impact of the parameter  $\rho$  on the *Capital-at-Risk*

## Loss Distribution Approach for operational risk

---

is equivalent to assume that the dependence function is the upper Fréchet copula  $\mathbf{C}^+$ . Let us consider two operational risks. Using results of FRANK, NELSEN and SCHWEIZER [1987], WILLIAMSON [1989] shows that the dependency bounds of  $\mathbf{G}$  when the dependence function is larger than a given copula  $\mathbf{C}_-$  are

$$\mathbf{G}_\vee(x) \leq \mathbf{G}(x) \leq \mathbf{G}_\wedge(x) \quad (84)$$

with

$$\mathbf{G}_\vee^{(-1)}(u) = \inf_{\mathbf{C}_-(u_1, u_2)=u} \mathbf{G}_1^{(-1)}(u_1) + \mathbf{G}_2^{(-1)}(u_2) \quad (85)$$

and

$$\mathbf{G}_\wedge^{(-1)}(u) = \sup_{\tilde{\mathbf{C}}_-(u_1, u_2)=u} \mathbf{G}_1^{(-1)}(u_1) + \mathbf{G}_2^{(-1)}(u_2) \quad (86)$$

where  $\tilde{\mathbf{C}}$  is the dual copula of  $\mathbf{C}$ . In the case where  $\mathbf{C}_- = \mathbf{C}^+$ , DURRLEMAN, NIKEGHBALI and RONCALLI [2000] shows that

$$\begin{aligned} \mathbf{G}^{(-1)}(u) &= \mathbf{G}_\vee^{(-1)}(u) \\ &= \mathbf{G}_\wedge^{(-1)}(u) \\ &= \inf_{\min(u_1, u_2)=u} \mathbf{F}_1^{(-1)}(u_1) + \mathbf{F}_2^{(-1)}(u_2) \\ &= \mathbf{G}_1^{(-1)}(u) + \mathbf{G}_2^{(-1)}(u) \end{aligned} \quad (87)$$

We deduce that<sup>24</sup>

$$\text{CaR}(\alpha) = \text{CaR}_1(\alpha) + \text{CaR}_2(\alpha) \quad (90)$$

Extension to multivariate case is straightforward. So, we have the following result

$$\mathbf{C} \langle \check{\mathbf{G}} \rangle = \mathbf{C}^+ \iff \text{CaR}(\alpha) = \sum_{i=1}^I \sum_{j=1}^J \text{CaR}(i, j; \alpha) \quad (91)$$

Note that  $\mathbf{C} \langle \check{\mathbf{G}} \rangle = \mathbf{C}^+$  implies that there exists a random variable  $\xi$  and  $I \times J$  non-decreasing functions  $\beta_{i,j}$  such that we have

$$\vartheta(i, j) = \beta_{i,j}(\xi) \quad (92)$$

$\mathbf{C}^+$  corresponds also to the case of totally positive dependence.

---

<sup>24</sup>We may consider an alternate proof by using the fact that  $\vartheta_2 = \mathbf{G}_2^{(-1)}(\mathbf{G}_1(\vartheta_1))$ . Let us denote  $\varpi$  the function  $x \mapsto x + \mathbf{G}_2^{(-1)}(\mathbf{G}_1(x))$ . We have

$$\begin{aligned} \alpha &= \Pr\{\vartheta_1 + \vartheta_2 \leq \text{CaR}(\alpha)\} \\ &= \mathbb{E}[\mathbf{1}_{[\varpi(\vartheta_1) \leq \text{CaR}(\alpha)]}] \\ &= \mathbf{G}_1(\varpi^{-1}(\text{CaR}(\alpha))) \end{aligned} \quad (88)$$

It comes that  $\text{CaR}(\alpha) = \varpi(\mathbf{G}_1^{(-1)}(\alpha))$  and we obtain the relationship

$$\text{CaR}(\alpha) = \mathbf{G}_1^{(-1)}(\alpha) + \mathbf{G}_2^{(-1)}(\mathbf{G}_1(\mathbf{G}_1^{(-1)}(\alpha))) = \text{CaR}_1(\alpha) + \text{CaR}_2(\alpha) \quad (89)$$

### B.2 Correlated frequencies

We focus now on the ‘correlation’ between frequencies of different types of risk

$$\mathbb{E}[N_{k_1}(t)N_{k_2}(t)] \neq \mathbb{E}[N_{k_1}(t)] \times \mathbb{E}[N_{k_2}(t)] \quad (93)$$

$N_k(t)$  is generally assumed to be a Poisson variable  $\mathcal{P}$  with mean  $\lambda_k$ . The idea is also to use a multivariate extension of the Poisson distribution. However, multivariate Poisson distributions are relatively complicated for dimensions higher than two (JOHNSON, KOTZ and BALAKRISHNAN [1997]). Let  $N_{11}$ ,  $N_{12}$  and  $N_{22}$  be three **independent** Poisson variates with means  $\lambda_{11}$ ,  $\lambda_{12}$  and  $\lambda_{22}$ . In the bivariate case, the joint distribution is based on the variables  $N_1 = N_{11} + N_{12}$  and  $N_2 = N_{22} + N_{12}$ . We have of course  $N_1 \sim \mathcal{P}(\lambda_1 = \lambda_{11} + \lambda_{12})$  and  $N_2 \sim \mathcal{P}(\lambda_2 = \lambda_{22} + \lambda_{12})$ . Moreover, the joint probability function is

$$\Pr\{N_1 = n_1, N_2 = n_2\} = \sum_{n=0}^{\min(n_1, n_2)} \frac{\lambda_{11}^{n_1-n} \lambda_{12}^{n_2-n} \lambda_{22}^n e^{-(\lambda_{11} + \lambda_{22} + \lambda_{12})}}{(n_1 - n)! (n_2 - n)! n!} \quad (94)$$

The Pearson correlation between  $N_1$  and  $N_2$  is  $\rho = \lambda_{12} [(\lambda_{11} + \lambda_{12})(\lambda_{22} + \lambda_{12})]^{-\frac{1}{2}}$  and it comes that

$$\rho \in \left[ 0, \min \left( \sqrt{\frac{\lambda_{11} + \lambda_{12}}{\lambda_{22} + \lambda_{12}}}, \sqrt{\frac{\lambda_{22} + \lambda_{12}}{\lambda_{11} + \lambda_{12}}} \right) \right] \quad (95)$$

With this construction, we have only positive dependence. In an operational risk context, it is equivalent to say that the two risks are affected by specific and systemic risks. Nevertheless, people in charge of operational risk in a bank have little experience with this approach and are more familiar with correlation concepts. To use this approach, it is also necessary to invert the previous relationships. In this case, we have

$$\begin{aligned} \lambda_{12} &= \rho \sqrt{\lambda_1 \lambda_2} \\ \lambda_{11} &= \lambda_1 - \rho \sqrt{\lambda_1 \lambda_2} \\ \lambda_{22} &= \lambda_2 - \rho \sqrt{\lambda_1 \lambda_2} \end{aligned} \quad (96)$$

In dimension  $K$ , there is a generalisation of the bivariate case by considering more than  $K$  independent Poisson variates. However, the corresponding multivariate Poisson distribution is not *tractable* because the correlation coefficients have not an easy expression.

SONG [2000] suggests then an interesting alternative by using a copula  $\mathbf{C}$ . In this case, the probability mass function is given by the Radon-Nikodym density of the distribution function:

$$\Pr\{N_1 = n_1, \dots, N_k = n_k, \dots, N_K = n_K\} = \sum_{i_1=1}^2 \dots \sum_{i_K=1}^2 (-1)^{i_1 + \dots + i_K} \mathbf{C} \left( \sum_{n=0}^{n_1} \frac{\lambda_1^{n+1-i_1} e^{-\lambda_1}}{n!}, \dots, \sum_{n=0}^{n_k} \frac{\lambda_k^{n+1-i_k} e^{-\lambda_k}}{n!}, \dots, \sum_{n=0}^{n_K} \frac{\lambda_K^{n+1-i_K} e^{-\lambda_K}}{n!} \right) \quad (97)$$

Assuming a Normal copula, we note  $\mathcal{P}(\boldsymbol{\lambda}, \boldsymbol{\rho})$  the multivariate Poisson distribution generated by the Normal copula with parameter  $\boldsymbol{\rho}$  and univariate Poisson distributions  $\mathcal{P}(\lambda_k)$  (to illustrate this distribution, we give an example in the following footnote<sup>25</sup>). We have to remark that the parameter of the Normal copula  $\boldsymbol{\rho}$  is not equal to the Pearson correlation matrix, but is generally very close (see the figure 27). The *Capital-at-Risk* with an  $\alpha$  confidence level could then be calculated by assuming that  $\mathbf{N} = \{N_1, \dots, N_k, \dots, N_K\}$  follows a multivariate Poisson distribution  $\mathcal{P}(\boldsymbol{\lambda}, \boldsymbol{\rho})$ . Moreover, there are no computational difficulties, because the estimation of the

---

<sup>25</sup>The next table contains the probability mass function  $p_{i,j} = \Pr\{N_1 = i, N_2 = j\}$  of the bivariate Poisson distribution  $\mathcal{P}(\lambda_1 = 1, \lambda_2 = 1, \rho = 0.5)$ .

parameters  $\lambda$  and  $\rho$  is straightforward, and the distribution can be easily obtained with Monte Carlo methods<sup>26</sup>.

## C Range of order statistics

Let us consider range of order statistics  $W_{m_1, m_2; n} = X_{m_2; n} - X_{m_1; n}$  ( $1 \leq m_1 < m_2 \leq n$ ). We can show that the density  $h_{W_{m_1, m_2; n}}(w)$  has the following expression<sup>27</sup>

$$h_{W_{m_1, m_2; n}}(w) = n! \int_{-\infty}^{\infty} \frac{[\mathbf{H}(x)]^{m_1-1}}{(m_1-1)!} \frac{[\mathbf{H}(x+w) - \mathbf{H}(x)]^{m_2-m_1-1}}{(m_2-m_1-1)!} \frac{[1 - \mathbf{H}(x+w)]^{n-m_2}}{(n-m_2)!} h(x) h(x+w) dx \quad (99)$$

For operational risk, the most interesting ranges are these of consecutive high order statistics  $W_{m:n}^\bullet = W_{m-1, m; n}$  (with  $m \simeq n$ ). It comes that the density of  $W_{m:n}^\bullet$  is (see BALAKRISHNAN and CLIFFORD COHEN [1991])

$$h_{W_{m:n}^\bullet}(w) = n! \int_{-\infty}^{\infty} \frac{[\mathbf{H}(x)]^{m-2}}{(m-2)!} \frac{[1 - \mathbf{H}(x+w)]^{n-m}}{(n-m)!} h(x) h(x+w) dx \quad (100)$$

For the distribution, we have<sup>28</sup>

$$H_{W_{m:n}^\bullet}(w) = 1 - n! \int_{-\infty}^{\infty} \frac{[\mathbf{H}(x)]^{m-2}}{(m-2)!} \frac{[1 - \mathbf{H}(x+w)]^{n-m+1}}{(n-m+1)!} h(x) dx \quad (103)$$

$p_{i,j}$	0	1	2	3	4	5	...	$p_{i,\cdot}$
0	0.0945	0.133	0.0885	0.0376	0.0114	0.00268		0.368
1	0.0336	0.1	0.113	0.0739	0.0326	0.0107		0.368
2	0.00637	0.0312	0.0523	0.0478	0.0286	0.0123		0.184
3	0.000795	0.00585	0.0137	0.0167	0.013	0.0071		0.0613
4	7.28E-005	0.000767	0.00241	0.00381	0.00373	0.00254		0.0153
5	5.21E-006	7.6E-005	0.000312	0.000625	0.000759	0.000629		0.00307
⋮								
$p_{\cdot,j}$	0.135	0.271	0.271	0.18	0.0902	0.0361		1

If  $\rho = -0.5$ , we obtain the following values for  $p_{i,j}$ .

$p_{i,j}$	0	1	2	3	4	5	...	$p_{i,\cdot}$
0	0.0136	0.0617	0.101	0.0929	0.058	0.027		0.368
1	0.0439	0.112	0.111	0.0649	0.026	0.00775		0.368
2	0.0441	0.0683	0.0458	0.0188	0.00548	0.00121		0.184
3	0.0234	0.0229	0.0109	0.00331	0.000733	0.000126		0.0613
4	0.00804	0.00505	0.00175	0.000407	7.06E-005	9.71E-006		0.0153
5	0.002	0.00081	0.000209	3.79E-005	5.26E-006	5.89E-007		0.00307
⋮								
$p_{\cdot,j}$	0.135	0.271	0.271	0.18	0.0902	0.0361		1

<sup>26</sup>The simulation of 20 millions draw of the loss distribution  $\mathbf{G}$  with 5 loss types takes less than one hour with a Pentium III 750 Mhz and the GAUSS programming language.

<sup>27</sup>We can derive this mathematical result by using the fact that the joint density of  $(X_{m_1; n}, X_{m_2; n})$  is

$$h_{m_1, m_2; n}(x_1, x_2) = n! \frac{[\mathbf{H}(x_1)]^{m_1-1}}{(m_1-1)!} \frac{[\mathbf{H}(x_2) - \mathbf{H}(x_1)]^{m_2-m_1-1}}{(m_2-m_1-1)!} \frac{[1 - \mathbf{H}(x_2)]^{n-m_2}}{(n-m_2)!} h(x_1) h(x_2) \quad (98)$$

<sup>28</sup>We use the following chain of equivalent statements:

$$\begin{aligned} H_{W_{m:n}^\bullet}(w) &= n! \int_0^w \int_{-\infty}^{\infty} \frac{[\mathbf{H}(x)]^{m-2}}{(m-2)!} \frac{[1 - \mathbf{H}(x+\omega)]^{n-m}}{(n-m)!} h(x) h(x+\omega) dx d\omega \\ &= n! \int_{-\infty}^{\infty} \frac{[\mathbf{H}(x)]^{m-2}}{(m-2)!} h(x) \left[ \frac{[1 - \mathbf{H}(x)]^{n-m+1}}{(n-m+1)!} - \frac{[1 - \mathbf{H}(x+w)]^{n-m+1}}{(n-m+1)!} \right] dx \\ &= 1 - n! \int_{-\infty}^{\infty} \frac{[\mathbf{H}(x)]^{m-2}}{(m-2)!} \frac{[1 - \mathbf{H}(x+w)]^{n-m+1}}{(n-m+1)!} h(x) dx \end{aligned} \quad (101)$$

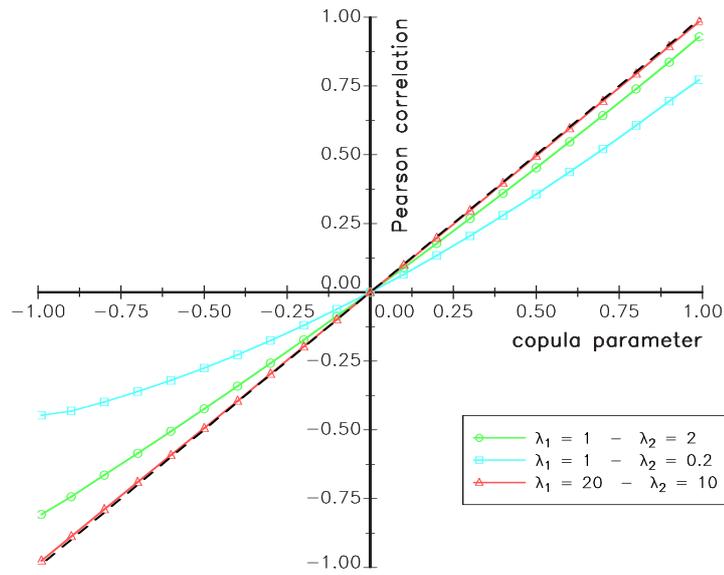


Figure 27: Relationship between the Normal copula parameter and the Pearson correlation

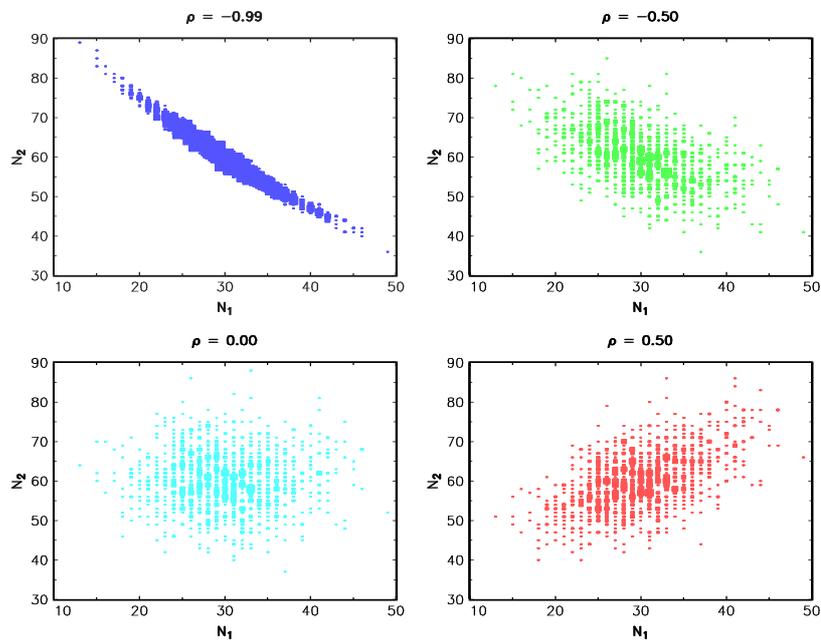


Figure 28: Random generation of bivariate Poisson variates  $\mathcal{P}(30)$  and  $\mathcal{P}(60)$

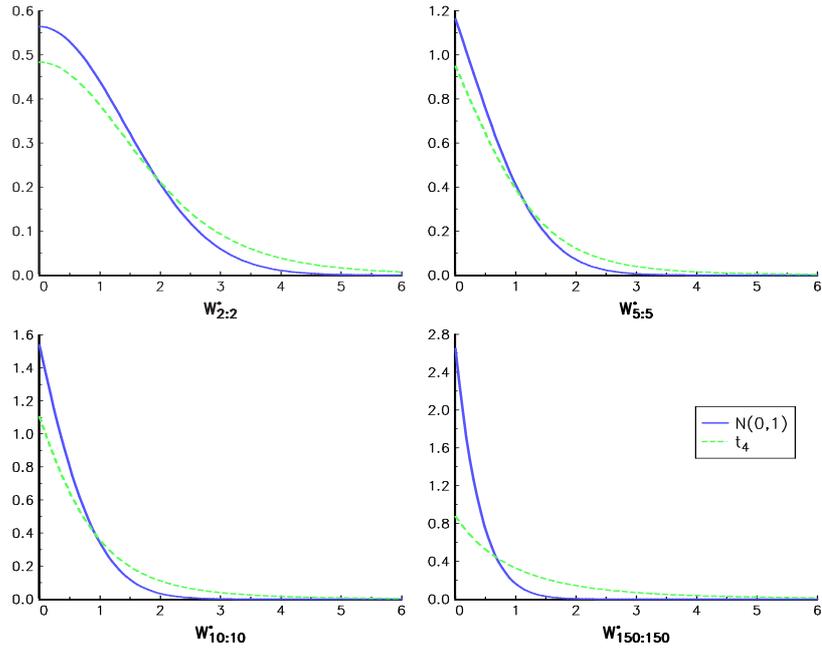


Figure 29: Impact of fat tails on the range statistic  $W_{n:n}^\bullet$

In practice, the integration are computed numerically using a Gauss-Hermite quadrature. As for order statistics, fat tails may have a big impact on the range statistics (see Figure 29). Because we have  $W_{m:n}^\bullet \simeq W_{n:n}^\bullet$  for  $m$  close to  $n$ , we can compare the distribution of  $W_{n:n}^\bullet$  with the observed statistics  $\{w_{m:n}^\bullet, m = n, n-1, n-2, \dots\}$ . In Figure 30, we have represented the survival function of  $W_{n:n}^\bullet$  for two distributions and the observed ranges  $\{w_{m:n}^\bullet, m = n, \dots, n-14\}$ . We remark that the first distribution can not explain the biggest ranges, whereas the second distribution can not explain the lowest ranges.

## D Estimation methods

### D.1 The method of maximum likelihood

Let  $\theta$  be the  $K \times 1$  vector of parameters to be estimated and  $\Theta$  the parameter space. The likelihood for observation  $t$ , that is the probability density of the observation  $t$ , considered as a function of  $\theta$ , is denoted  $L_t(\theta)$ . Let  $\ell_t(\theta)$  be the log-likelihood of  $L_t(\theta)$ . Given  $T$  observations, we get

$$\ell(\theta) = \sum_{t=1}^T \ell_t(\theta) \quad (104)$$

In the last calculus, we use the fact that

$$\begin{aligned} \int_{-\infty}^{\infty} [\mathbf{H}(x)]^h [1 - \mathbf{H}(x)]^k h(x) dx &= \left[ \frac{[\mathbf{H}(x)]^{h+1}}{h+1} [1 - \mathbf{H}(x)]^k \right]_{-\infty}^{\infty} + n \int_{-\infty}^{\infty} \frac{[\mathbf{H}(x)]^{h+1}}{h+1} [1 - \mathbf{H}(x)]^{k-1} h(x) dx \\ &= \frac{h!k!}{(h+k+1)!} \end{aligned} \quad (102)$$

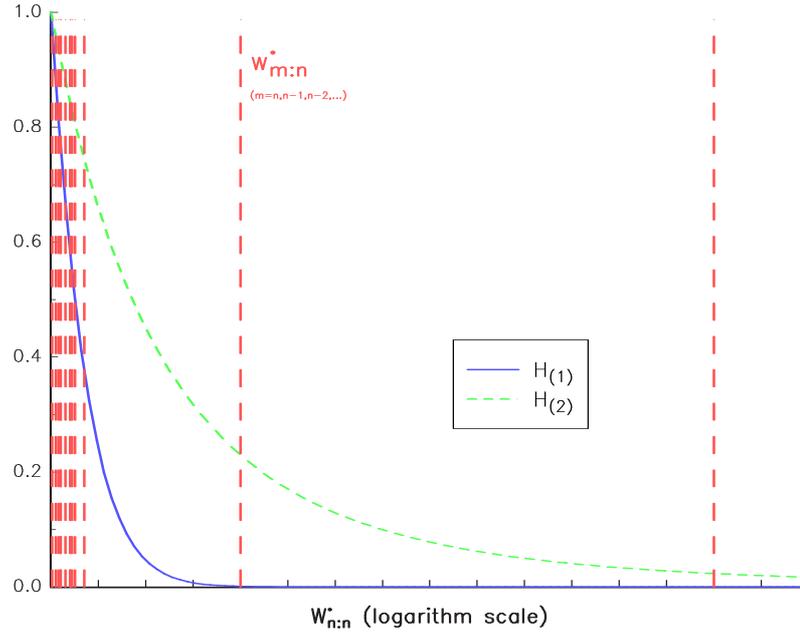


Figure 30: Survival function of  $W_{m:n}^{\bullet}$  ( $m = n, n - 1, n - 2, \dots$ ) and observed ranges

the log-likelihood function.  $\hat{\theta}_{\text{ML}}$  is the Maximum Likelihood estimator if

$$\ell(\hat{\theta}_{\text{ML}}) \geq \ell(\theta) \quad \forall \theta \in \Theta \quad (105)$$

We may show that  $\hat{\theta}_{\text{ML}}$  has the property of **asymptotic normality** (DAVIDSON and MACKINNON [1993]) and we have

$$\sqrt{T}(\hat{\theta}_{\text{ML}} - \theta_0) \longrightarrow \mathcal{N}(\mathbf{0}, \mathcal{J}^{-1}(\theta_0)) \quad (106)$$

with  $\mathcal{J}(\theta_0)$  the **Fisher information matrix**. Let  $J_{\hat{\theta}_{\text{ML}}}$  be the  $T \times K$  Jacobian matrix of  $\ell_t(\theta)$  and  $H_{\hat{\theta}_{\text{ML}}}$  the  $K \times K$  Hessian matrix of the likelihood function<sup>29</sup>. The covariance matrix of  $\hat{\theta}_{\text{ML}}$  in finite sample could be estimated by the inverse of the negative Hessian

$$\text{var}(\hat{\theta}_{\text{ML}}) = (-H_{\hat{\theta}_{\text{ML}}})^{-1} \quad (108)$$

or by the inverse of the OPG estimator

$$\text{var}(\hat{\theta}_{\text{ML}}) = (J_{\hat{\theta}_{\text{ML}}}^{\top} J_{\hat{\theta}_{\text{ML}}})^{-1} \quad (109)$$

There is another one called the White (or “sandwich”) estimator defined by

$$\text{var}(\hat{\theta}_{\text{ML}}) = (-H_{\hat{\theta}_{\text{ML}}})^{-1} (J_{\hat{\theta}_{\text{ML}}}^{\top} J_{\hat{\theta}_{\text{ML}}}) (-H_{\hat{\theta}_{\text{ML}}})^{-1} \quad (110)$$

---

<sup>29</sup>These different matrices are defined by

$$\begin{aligned} J_{\hat{\theta}_{\text{ML}}} &= \left[ \frac{\partial \ell_t}{\partial \theta^{\top}} \right]_{\theta=\hat{\theta}_{\text{ML}}} \\ H_{\hat{\theta}_{\text{ML}}} &= \left[ \frac{\partial^2 \ell}{\partial \theta \partial \theta^{\top}} \right]_{\theta=\hat{\theta}_{\text{ML}}} \end{aligned} \quad (107)$$

which takes into account heteroskedasticity.

### D.2 Generalized method of moments

We consider that the empirical moments  $h_{t,m}(\theta)$  depend on the  $K \times 1$  vector  $\theta$  of parameters.  $T$  is the number of observations and  $M$  is the number of conditions on the moments. Let  $\theta_0$  be the “true” value of  $\theta$ . The generalized method of moments is based on the assumption that

$$\mathbb{E}[h_{t,m}(\theta_0)] = 0 \quad (111)$$

Consider  $\mathbf{h}_t(\theta)$  the row vector of the elements  $h_{t,1}(\theta), \dots, h_{t,M}(\theta)$  and  $H(\theta)$  the  $T \times M$  matrix with elements  $h_{t,m}(\theta)$ . Let  $g(\theta)$  be a  $M \times 1$  vector given by

$$g_m(\theta) = \frac{1}{T} \sum_{t=1}^T h_{t,m}(\theta) \quad (112)$$

The GMM criterion function  $Q(\theta)$  is defined by:

$$Q(\theta) = g(\theta)^\top W^{-1} g(\theta) \quad (113)$$

with  $W$  a symmetric positive definite  $M \times M$  matrix. The GMM estimator  $\hat{\theta}_{\text{GMM}}$  corresponds to

$$\hat{\theta}_{\text{GMM}} = \arg \min_{\theta \in \Theta} Q(\theta) \quad (114)$$

Like the ML estimates, we may show that  $\hat{\theta}_{\text{GMM}}$  has the property of **asymptotic normality** and we have

$$\sqrt{T} \left( \hat{\theta}_{\text{GMM}} - \theta_0 \right) \longrightarrow \mathcal{N}(\mathbf{0}, \Sigma) \quad (115)$$

In the case of **optimal weights** ( $W$  is the covariance matrix  $\Phi$  of  $H(\theta)$  – HANSEN [1982]), we have

$$\text{var} \left( \hat{\theta}_{\text{GMM}} \right) = \frac{1}{T} \left[ D^\top \hat{\Phi}^{-1} D \right]^{-1} \quad (116)$$

with  $D$  the  $M \times K$  Jacobian matrix of  $g(\theta)$  computed for the estimate  $\hat{\theta}_{\text{GMM}}$ .

“The ML estimator can be treated as a GMM estimator in which empirical moments are the components of the score vector” (DAVIDSON and MACKINNON [1993]). ML method is also a special case of GMM with  $g(\theta) = \partial_\theta \ell(\theta)$  and  $W = I_K$ . That’s why  $\text{var} \left( \hat{\theta}_{\text{GMM}} \right)$  is interpreted as an OPG estimator.

### D.3 Simulated method of moments

To implement GMM, we have to define the *orthogonal conditions*. So, we need analytical expressions for the moment conditions. When analytical expressions are not available, DUFFIE and SINGLETON [1993] suggest to use simulated moment conditions. This method is very easy to implement and the only thing that differs from GMM is the computation of the covariance matrix.

### D.4 Indirect inference estimation

Indirect inference methods have been first developed by GOURIÉROUX, MONFORT and RENAULT [1993] and GALLANT and TAUCHEN [1996]. These methods are particularly suited for our case where the direct likelihood function cannot be computed easily. As said previously, the statistical treatment of operational risk requires to calibrate the probability distribution of sums of independently distributed random variables, especially the only available data are often partially aggregated. Unfortunately, except if the probability distribution of these variables is assumed to be gaussian, we can never write the likelihood function in a closed-form way. As a result, performing a maximisation of the likelihood function is hardly feasible.

The idea of the indirect inference method is rather simple. It is based on two distribution functions for the aggregated loss distribution:

## Loss Distribution Approach for operational risk

---

- the “true” one denoted by  $\mathbf{F}(\xi, \theta)$  where  $\theta$  is a set of parameters to be calibrated. As an example, if we assume that individual losses follow a Log-Normal distribution,  $\theta$  stands for the mean and the standard deviation of the logarithm of the loss;
- an approximate distribution denoted by  $\mathbf{F}^*(\xi, \eta)$  where  $\eta$  is also a set of parameters.

Furthermore, we consider the case where:

- maximizing the true likelihood — based on  $\mathbf{F}(\xi, \theta)$  — is not feasible;
- however, one can always simulate data following this true distribution;
- maximizing the approximate distribution is always a tractable task.

This is typically the case when the distribution of  $\zeta$  is Log-Normal and the approximate one of  $\xi$  is Normal. If we denote  $\boldsymbol{\xi} = (\xi_1, \dots, \xi_T)$  the set of observations (as said, some of these observations are aggregated losses and the others are individual losses), the algorithm is then the following one:

1. Perform a maximum likelihood scheme **as if** these losses  $\xi_t$  followed the approximate distribution. Maximize the approximate log-likelihood:

$$\hat{\eta} = \arg \max \ell^*(\boldsymbol{\xi}, \eta)$$

where

$$\ell^*(\boldsymbol{\xi}, \eta) = \sum_{t=1}^T \ln f^*(\xi_t, \eta) \tag{117}$$

2. Take a starting value  $\theta$ . Simulate a set of data  $\boldsymbol{\xi}^{(s)}$  drawn from the true distribution  $\mathbf{F}(\xi, \theta)$  and maximize the approximate log-likelihood

$$\hat{\eta}^{(s)} = \arg \max \sum_{t=1}^T \ln f^*(\xi_t^{(s)}, \eta^{(s)}) \tag{118}$$

3. Compute

$$\tilde{\eta}(\theta) = \frac{1}{S} \sum_{s=1}^S \hat{\eta}^{(s)} \tag{119}$$

4. Iterate steps 2 and 3 until obtaining the solution  $\hat{\theta}$  of  $\tilde{\eta}(\theta) = \hat{\eta}$ .

It can be shown that  $\hat{\theta}$  is a consistent estimator of the true parameter when  $S$  grows to infinity.

Some remarks are worth noting. First, this algorithm can be greatly simplified regarding its computational burden. Indeed, steps 2 and 3 need a maximization at each iteration. This can be overcome using Gallant and Tauchen approach. Secondly, the closer  $\mathbf{F}$  and  $\mathbf{F}^*$  are, the better  $\hat{\theta}$  is. As a result, one may benefit from using an approximate distribution which is not too far away from the true distribution. Thirdly, this method can be seen as a variant of the simulated method of moments.