

# Computational Finance and its Applications

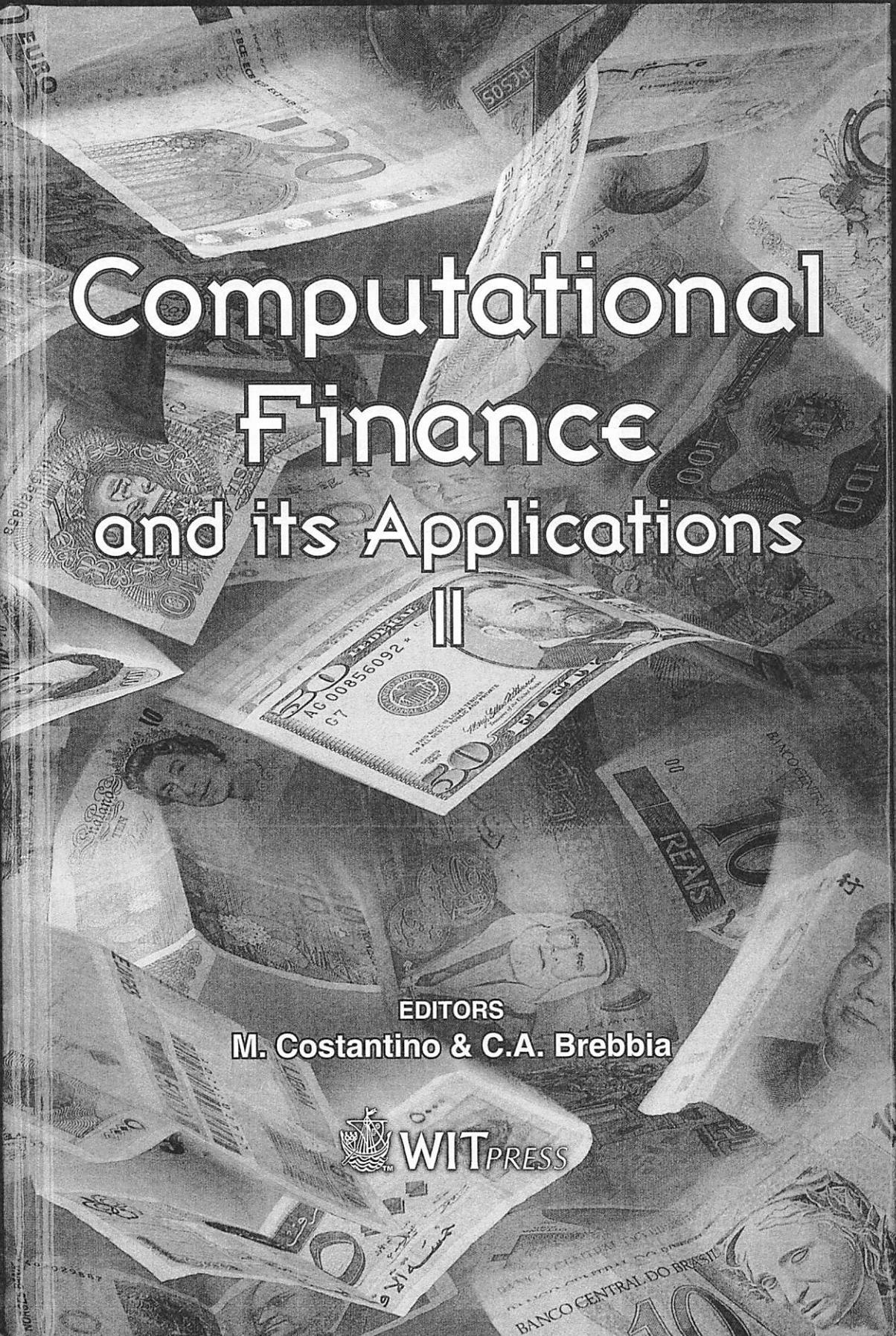
II

EDITORS

M. Costantino & C.A. Brebbia



**WIT**PRESS



## Section 7

# Risk management

## Monte Carlo risk management

M. Di Pierro & A. Nandy

*CTI, DePaul University, Chicago, IL, USA*

### Abstract

In this paper we propose a Monte Carlo based approach to Risk Management. Our approach applies to any system subject to random uncorrelated losses under very general conditions. Our methodology consists of the following steps: model the distribution of losses and the distribution of time intervals between losses via an analysis of historical data; perform a Monte Carlo simulation of a finite period of time in the future; use the Monte Carlo data to estimate the 99.9% VaR.

### 1 Introduction

In this paper we present an approach to Risk Management based on the Monte Carlo technique. The proposed approach is very general and it can be applied to any system characterized by discrete losses. We have made the following broad assumptions: a) loss events are independent; b) number of loss events occurring in any time interval  $\Delta T$  is independent of loss events occurring before the time interval considered; c) the probability of two events occurring at exactly same time is zero.

Discrete losses due to internal fraud (i.e. Operational Risk as defined by Basel II accord) for a particular department in a Bank provide a good example of application [1].

Our approach is based on the following steps: 1) model the time distribution and the severity distribution of losses; 2) simulate possible future scenarios compatible with the observed time and severity distributions; 3) compute the 99.9% VaR as the monetary amount that is greater than the total loss in 99.9% of the simulated scenarios.

Notice that our approach is free from bias or assumptions about the distribution of the total loss. The validity of the assumptions a), b) and c) can be verified directly from the data, as we show in the example of section 3.



In the next section of this paper we present the general approach and provide examples of distribution functions that can be used to model the time distribution and the severity distribution of losses.

In the third section we provide an example of application. We build a hypothetical portfolio of stocks and compute the 99.9% VaR due exclusively to mini-crashes. In the example we define a mini-crash as the event when all stocks considered move simultaneously downward. We stress again that our method is very general and can be applied seamlessly to any portfolio for any quantifiable definition of crashes [2] (as long as they can be identified as discrete events), as well as to the analysis of other types of losses (for example Operational Risk losses).

We want to stress that in the cases examined here, usual analytical formulas for computing the VaR without performing a simulation does not apply because the distribution of the total loss over a finite period of time is not Gaussian.

## 2 Model

Our approach is based on the following assumptions:

- a) loss events are independent.
- b) the number of loss events occurring in any time interval  $\Delta T$  is independent of the number of loss events occurring before the time interval considered;
- c) the probability of two events occurring at exactly same time is zero.

The consequence of assumptions a), b) and c) is that we can separately model the frequency distribution of losses and the severity distribution of losses. More specifically, assumptions b) and c) indicate that the process underlying the frequency distribution of losses is a Poisson process.

With the above assumptions, our approach requires the following steps that are described in detail in the following subsections:

1. Model the time distribution and severity distribution of losses separately;
2. Simulate a large set of possible future scenarios compatible with the observed time and severity distributions;
3. Compute the 99.9% VaR as the monetary amount that is greater than the total loss in 99.9% of the simulated scenarios.

### 2.1 Modeling time and severity distributions

Each loss event is characterized by the time when the event occurs,  $T_i$ , and by the severity of the loss,  $S_i$ . In the example below we assume  $T_i$  is measured in days and  $S_i$  is measured in dollars. We also assume that losses are sorted by time so that  $T_i < T_{i+1}$ . We model the time distribution of  $\{T_i\}$  and  $\{S_i\}$  separately.

As consequence of assumptions a), b) and c), the process underlying the frequency distribution of losses is a Poisson process. This means that the number of events occurring in any finite time frame  $T$  follows the Poisson distribution. It



also means that the distribution of the time intervals between subsequent losses,  $t_i = T_{i+1} - T_i$  is exponential

$$p(t) = \lambda e^{-\lambda t} \quad (1)$$

where  $\lambda$  is one of the parameter of our model and it equals the average number of events per unit of time (per day). It can be measured from the data as

$$\lambda = (E[t])^{-1} = \left( \frac{1}{N-1} \sum_{i=0}^{i < N-1} (T_{i+1} - T_i) \right)^{-1} \quad (2)$$

where  $N$  is the number of historical loss events.

Modeling the severity distribution is a more difficult task and it depends on the data set at hand. Some general characteristics are that losses are always positive and can be arbitrarily large or small, therefore we expect the underlying distribution to be appreciatively log-normal. In the most general case, consistently with extreme value theory [3–5], we expect the presence of a fat tail for high losses, as observed in [1, 6]. A fat tail is characterized by an inverse polynomial behavior as opposed to an exponential behavior as for a normal or lognormal distribution. The natural choice for modeling this fat tail is therefore a generalized Pareto distribution (GPD).

Our approach consists of modeling the severity distribution using the following distribution function

$$p(x) = a\theta(x - x_0) \frac{1}{x} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}} + ab\theta(x_0 - x)(x + c)^{-\beta} \quad (3)$$

which is discussed in detail in Appendix A. This distribution is exactly log-normal for  $x < x_0$ , exactly GPD for  $x > x_0$ , and it is continuous and differentiable everywhere. It reduces to a log-normal for  $x_0 \rightarrow \infty$  and to a GPD for  $x_0 = 0$ . It depends on only four parameters:  $\mu$ ,  $\sigma$  (mean and variance of the log-normal contribution),  $x_0$  (the point where change of behavior is observed),  $\beta$  (the power of the polynomial behavior of the fat tail).  $a$ ,  $b$ , and  $c$  are constants that depend on the above four parameters and their explicit analytical expressions are reported in Appendix A.

Fig. (1) shows various plots of  $p(x)$  (left) and the corresponding cumulative distribution functions  $F(x)$  (right) in logarithmic scale, for different values of  $x_0$ , and fixed values of  $\mu$ ,  $\sigma$ , and  $\beta$ . The greater the value of  $x_0$ , the smaller the fat tail.

In summary, the parameters of our model are  $\lambda$ ,  $\mu$ ,  $\sigma$ ,  $x_0$ , and  $\beta$ . While  $\lambda$  is measured via eq. (2), the other 4 parameters can be measure via a fit of the cumulative distribution function implied by historical severity data.

## 2.2 Simulation

Once the model parameters are extracted from historical data, we proceed by performing a simulation of multiple future scenarios. Each simulated scenario



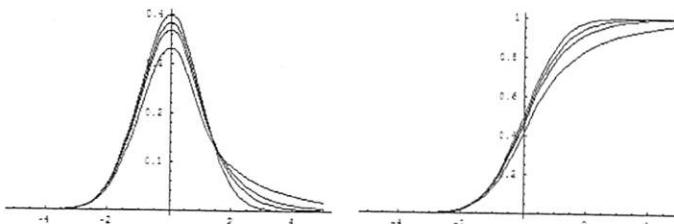


Figure 1: Log-plot of  $p(x)$  (left) and  $F(x) = \int_{-\infty}^x p(x)dx$  (right) for different values of  $\beta = 1.5, 2, 2.5$  at  $\mu = 0$ ,  $\sigma = 1$ , and  $z_0 = 1$  fixed, compared with a Gaussian distribution.

covers a finite time interval from 0 (today) until some future time  $T$ . This is the time over which we wish to compute the Value-at-Risk. Each simulated scenario is characterized by a set of simulated loss events occurring within the time interval  $T$ .

The simulation of each simulated scenario starts at time 0 and recursively proceeds to simulate the next loss event. The next loss event will occur in time  $t$  and will have a severity  $s$ .  $t$  and  $s$  are random numbers generated according to the frequency distribution and the severity distribution respectively. The simulation of each scenario proceeds until a loss event occurs beyond the time interval  $T$  being considered.

Multiple scenarios are simulated using the method described above. Then, for each simulated scenario  $k$ , we compute the total loss  $L_k$  by summing the discounted  $s$  losses occurring under scenario  $k$  within the considered time interval.

### 2.3 Computing the VaR

The 99.9% Value-at-Risk is defined as the value such that the probability of losing more than its value is 0.1%. Since our simulated scenarios are generated using the same procedure, they all occur with equal probability  $1/m$ , where  $m$  is the number of simulated scenarios. The 99.9% VaR can therefore be determined as that value that is greater than the total loss  $L_k$  in 99.9% of the simulated scenarios, and less the total loss in the remaining 0.1% of the simulated scenarios. This computation is done by sorting the scenarios according to their total loss  $L_k$  so that  $L_{k+1} > L_k$  and choosing (the notation  $\lceil x \rceil$  indicates excess rounding of  $x$ )

$$99.9\% \text{ VaR} = L_{\lceil 0.999m \rceil}$$

The same argument applies to the computation of the VaR for any other percentile. The larger is the number of simulated scenarios  $m$ , the more precise is the determination of the VaR.

Notice how, in general, the usual formula to compute the VaR in terms of the standard deviation of the losses does not apply. In particular that formula does not apply if the distribution of the total loss is not Gaussian. Fig. (2) shows that this is clearly not the case for the systems under consideration in this paper.



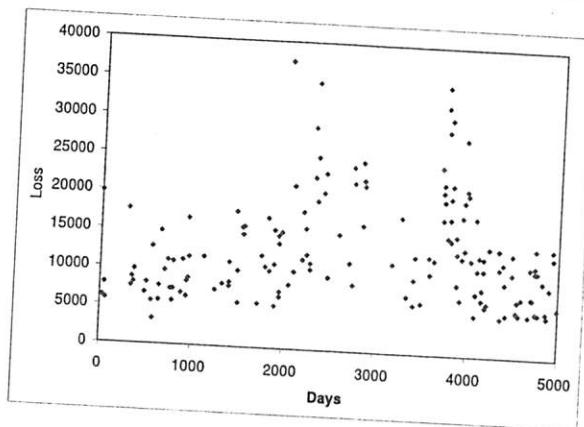


Figure 2: Loss events associated to the mini-crashes (all stocks moving downward) of our sample portfolio during the time frame March 16, 2001 - February 21, 2006.

### 3 Example

In this section, as an example application of our approach, we consider the computation of VaR due to mini-crashes in a portfolio of stocks. This number represents the economic capital that one would have to save to provide insurance against such events at 99.9% confidence level. We define a mini-crash as the event when all the stocks considered move simultaneously downward.

In order to provide a concrete case we consider a portfolio comprised of the following stocks:

MSFT, RTN, HCA, EP, AVP, HNZ, USB, GD

with a constant capital of \$100,000 invested in each of them.

The above set of stocks have been chosen because of their relative low correlation. If the stocks were uncorrected, the probability of all of them moving downward (a mini-crash) in one day would be about 0.4%. Our analysis of historical market data indicates that this occurred 159 times in the time period starting March 16, 1992 and ending February 21, 2006, which averages to approximately once every 31 calendar days, or once every 21 trading days. This corresponds to a mini-crash probability of 5%, ten times more likely than the naive expectation.

Fig. 2 shows past losses associated with our portfolio in the time frame considered.

Fig. 3 shows the distribution of time intervals between two consecutive past losses, superimposed with the exponential distribution using the value  $\lambda = 1/31$  extracted from the data. The plot indicates that the probability of two consecutive mini-crashes occurring in two consecutive days is more likely than predicted by the exponential distribution. This is a small deviation from the assumption and



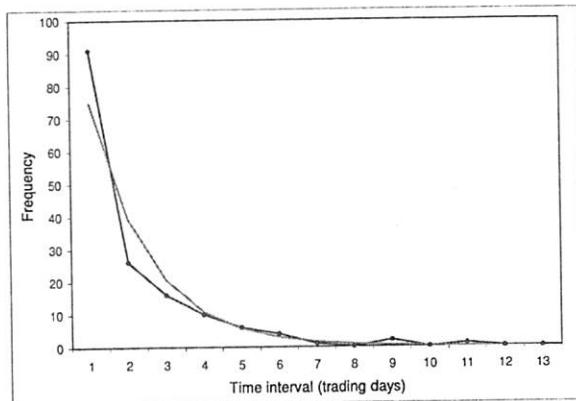


Figure 3: Distribution of time intervals between two consecutive past losses, superimposed to the exponential distribution at  $\lambda = 1/31$ .

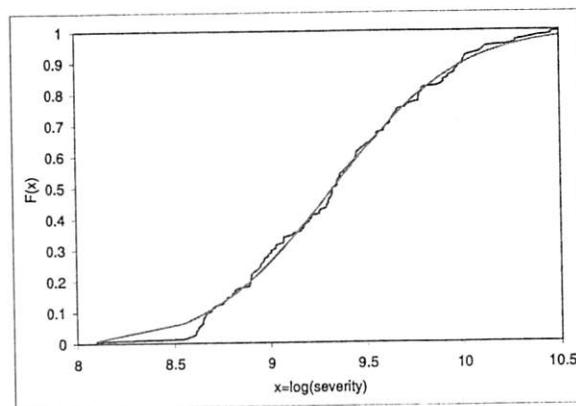


Figure 4: Log-plot of the cumulative distribution function  $F(x)$  for the severity of historical losses, and best fit using the distribution in eq. (3).

therefore it will be ignored in the rest of the paper, although it could be taken into account by correcting the frequency distribution.

Fig. 4 shows, in logarithmic scale, the cumulative distribution function for the severity of historical losses and its best fit using the distribution in eq. (3).

Fig. 5 shows the total loss for each simulated scenario for  $m = 10000$  simulations. The simulated scenarios are sorted according to their total loss. The 99.9% VaR reads directly from this plot as the  $y$ -axis value corresponding to the  $x$ -axis value  $[0.999m] = 9990$ .

Other VaR percentiles can be calculated in a similar fashion.

Our result is 99.9% VaR = \$360,000. This is 45% of the total funds invested in the portfolio. (The number exceeds the regular 99.9% VaR associated with the portfolio because it only consider events when all stocks jump downward. In fact,



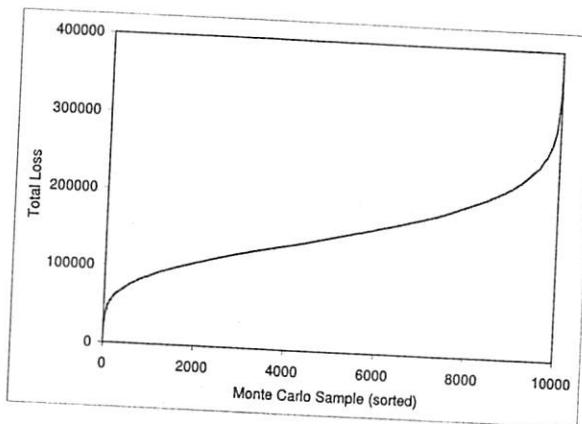


Figure 5: Total loss of each of the 10000 simulated scenarios. The 99.9% VaR reads directly from this plot as the  $y$ -axis value corresponding to the  $x$ -axis value  $[0.999m] = 9990$ .

in the example considered, positive jumps offset most of the risk discussed in this paper. Our approach allows us to isolate the risk associate to this specific type of discrete event.)

#### 4 Conclusion

In this paper we present a Monte Carlo approach to the computation of Value-at-Risk for systems subject to discrete losses. The main ingredients of our approach are that it model the frequency distribution of losses as a Poisson process and the severity distribution of losses using the novel fat-tail distribution discussed in this paper. The VaR is computed by performing a Monte Carlo simulation of future losses and sorting these scenarios according to their relative total loss. This procedure is very general and does not require any assumption about the distribution of the total loss.

In this paper, as an example, we have applied our technique to the computation of the VaR due to mini-crashes of a given stock portfolio.

#### Acknowledgements

We thank MetaCryption LLC for providing access to the financial data and access to the Monte Carlo Engine [7] used for our simulation.

#### Appendix A: Combined Log-normal + Pareto distribution

The particular distribution utilized in this paper to model the severity of losses has the following form



$$p(x) = a\theta(x - x_0) \frac{1}{x} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}} + ab\theta(x_0 - x)(x + c)^{-\beta} \quad (4)$$

This distribution is exactly log-normal for  $x < x_0$ , exactly Pareto for  $x > x_0$ , and it is continuous and differentiable everywhere, including  $x = x_0$ .

$\mu, \sigma, x_0$  and  $\beta$  are free parameters, while  $a, b$  and  $c$  are coefficients that depend on those parameters via the following exact relations:

$$a = \left( \sqrt{\frac{\pi}{2}} \sigma \left( 1 + \operatorname{erf}\left(\frac{\log x_0 - \mu}{\sqrt{2}\sigma}\right) \right) + \frac{b}{\beta - 1} (x_0 + c)^{1-\beta} \right)^{-1} \quad (5)$$

$$b = \frac{(x_0 + c)^\beta}{x_0} e^{-\frac{(\log x_0 - \mu)^2}{2\sigma^2}} \quad (6)$$

$$c = \frac{x_0 \beta \sigma^2}{\sigma^2 + \log x_0 - \mu} - x_0 \quad (7)$$

where we defined

$$\operatorname{erf}z = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt \quad (8)$$

## Appendix B: Statistics about the sample portfolio

Average daily log-return and daily volatility in the time period starting March 16, 1992 and ending February 21, 2006.

	MSFT	RTN	HCA	EP	AVP	HNZ	USB	GD
$\mu/100$	7.2%	1.6%	3.8%	3.1%	6.3%	3.6%	5.6%	11.4%
$\sigma$	2.2%	2.1%	2.2%	2.0%	2.0%	1.4%	1.9%	1.8%

Average correlation of daily log-returns in the same period.

	MSFT	RTN	HCA	EP	AVP	HNZ	USB	GD
MSFT	-	8%	13%	13%	10%	12%	24%	16%
RTN	8%	-	13%	10%	14%	13%	17%	29%
HCA	13%	13%	-	9%	17%	18%	21%	14%
EP	13%	10%	9%	-	7%	12%	15%	12%
AVP	10%	14%	17%	7%	-	24%	20%	14%
HNZ	12%	13%	18%	12%	24%	-	22%	14%
USB	24%	17%	21%	15%	20%	22%	-	17%
GD	16%	29%	14%	12%	14%	14%	17%	-

## References

- [1] M. Di Pierro and A. Nandy, "Comprehensive Modeling of Operational Risk Data An Empirical Approach", (publication-pending)
- [2] M Rubinstein, "Portfolio Insurance and the Market Crash", Financial Analysts Journal, (1988)
- [3] J. Beirlant, J. Teugels, P. Vynckier, P. "Practical analysis of extreme values", Leuven University Press, Leuven (1996)
- [4] P. Embrechts, C. Klüppelberg,T. Mikosch, "Modeling extremal events for insurance and Finance, Springer, Berlin (1997)
- [5] A. Roehr, "Modelling Operational Losses", Algo Research Quarterly, Vol. 5, n.2 (2002)
- [6] de Fontnouvelle, Patrick, De Jesus-Rueff, Virginia, Jordan, John S. and Rosengren, Eric S., "Using Loss Data to Quantify Operational Risk" (April 2003). <http://ssrn.com/abstract=395083>
- [7] OpRisk Calculator, [www.metacryption.com](http://www.metacryption.com) (a software tool developed by MetaCryption LLC that implements the methodology described in this paper).

