An efficient threshold choice for operational risk capital computation

Dominique Guegan, Bertrand Hassani, Cédric Naud

Documents de Travail du
Centre d'Economie de la Sorbonne

CES Working Papers

An efficient threshold choice for operational risk capital computation

Dominique GUEGAN, Bertrand HASSANI, Cédric NAUD

2010.96

Version révisée

Maison des Sciences Économiques, 106-112 boulevard de L'Hôpital, 75647 Paris Cedex 13
http://centredeconomiesorbonne.univ-paris1.fr/bandeau-haut/documents-de-travail/
ISSN : 1955-611X

An efficient threshold choice for operational risk capital computation.

November 1, 2011

Authors

- Dominique Guégan: Université Paris 1 Panthéon-Sorbonne, CES UMR 8174, 106 boulevard de l'Hopital 75647 Paris Cedex 13, France, e-mail: dguegan@univ-paris1.fr
- Bertrand K. Hassani: BPCE and Université Paris 1 Panthéon-Sorbonne, CES-MSE 106 boulevard de l'Hopital 75647 Paris Cedex 13, France, phone: +44 (0) 2070860973 , e-mail: bertrand.hassani@malix.univ-paris1.fr
- Cédric Naud: AON, 31-35 rue de la Fédération 75717 Paris CEDEX 15, phone: + 33 (0) 147831010, e-mail: cedric.naud@aon.fr

Documents de Travail du Centre d'Economie de la Sorbonne - 2010.96 (Version révisée)

Abstract

Operational risk quantification requires dealing with data sets which often present extreme values which have a tremendous impact on capital computations (VaR). In order to take into account these effects we use extreme value distributions, and propose a two pattern model to characterize loss distribution functions associated to operational risks: a lognormal on the corpus of the severity distribution and a Generalized Pareto Distribution on the right tail. The threshold from which the model switches form a scheme to the other one is obtained using a bootstrap method. We use an extension of the Peak-over-threshold method to fit the GPD and the EM algorithm to estimate the lognormal distribution parameters. Through the VaR, we show the impact of the GPD estimation procedure on the capital requirements. Besides, our work points out the importance of the building's choice of the information set by practitioners to compute capital requirements and we exhibit some incoherences with the actual rules. Particularly, we highlight a problem arising from the granularity which has recently been mentioned by the Basel Committee for Banking Supervision.

Keywords: Operational risk - Generalized Pareto distribution - Pickands estimate - Hill estimate - Expectation Maximization algorithm - Monte Carlo simulations - VaR.

1 Introduction

The purpose of this paper is operational risks quantification following the Basel Advanced Measurement Approach (AMA) (BIS (2004)). This proposal of the Basel committee binds banks to carry out their own models on internal data sets to evaluate amounts of capital necessary to face these risks.

In this paper, we focus on the computation of the loss distribution function (LDF) (Frachot et al. (2001), Guégan and Hassani (2009)) which is currently used to evaluate this kind of risks (Cruz (2004), Chernobai et al. (2007) and Shevchenko (2011)). We compute it as a convolution of two distribution functions, one modeling the frequencies of the losses and the other one their severity. Following the literature, we restrict to a Poisson distribution for the frequencies, and in this paper we focus on the estimation of the severity distribution.

The objective of the paper is to model the severities using a class of distributions which take into account the large losses: the Generalized Pareto Distribution (GPD). In order to capture extreme values on the first hand, and on the other hand the multimodality observed in operational risk data sets[1], we propose to model the severities as a mix of two distributions: one characterizing the most important losses denoted *cauda* using the GPD (Pickands (1975), Davis and Resnick (1988) and Embrechts et al. (1997)), and another one on the remaining data corresponding to the central part of the distribution or *corpus*. A characteristic of the GPD stands in the determination of a threshold denoted u permitting to define the data set on which the distribution will be adjusted. Therefore, the estimation of this threshold is the core of this paper: we highlight the influence of the choice of the threshold on the computation of the capital requirement. Then, once the threshold has been identified we need to estimate the parameters of the two distributions, for instance the lognormal and the GPD. Using several estimation methods, we also exhibit the influence of these ones on the computation of univariate capital requirement. Finally, considering the obtained results, we discuss the problem of granularity which is the main problem in the construction of operational risk data sets (or Basel matrix).

[1] We used data from BPCE on the French Caisse d'Epargne perimeter.

Documents de Travail du Centre d'Economie de la Sorbonne - 2010.96 (Version révisée)

The paper is organized as follows: Section two introduces the data set, Section three presents the theoretical framework, Section four describes the estimation of the threshold and the construction of the LDF and Section five is devoted to capital charge estimations. A sixth Section concludes.

2 The Data Set

The data set we use is organized into the Basel Matrix (BIS (2001)). In its first level of granularity, this matrix is made up of 56 cases - 8 business lines ("b") $\times$ 7 event types ("e")[2]. Nevertheless, each event type might be decomposed in several elements. For example, the "external fraud" event may be shared in two items - "Theft and Fraud" and "Systems Security" (second level of granularity). In a third level, the element "Theft and Fraud" may be split in several components: "Theft/Robbery", "Forgery" and "Check kiting". After a deep analysis, we observe that the kind of losses expected from a fraud with a credit card does not correspond to losses caused by someone hacking the system for instance; nevertheless they are in the same cell. Therefore, considering the largest level of granularity, we could face multimodal empirical distributions. Consequently, the methods used to model the losses depend on the granularity level choice. This choice might have a tremendous impact on capital requirement computations. Besides, we face a trade-off between quantity of data and robustness of the estimations: indeed, if the quantity of data is not sufficient, we cannot go lower in the granularity; on the another hand the ensuing empirical distribution is therefore an aggregate of various natures of data and the estimation of this last empirical distribution can be source of unusable or unrealistic results.

The previous paragraph dealt with the multimodality and the granularity problems. These points have a corollary: an extreme value is a relative concept, which require a referent. For example, a card fraud of 50000€ is an extreme value compared to an average loss of 300€ while a rogue trading of 50000€ is not an extreme value when we know that the average loss represents several millions. Furthermore, we do not define an extreme value as an "extremely" rare event. Therefore, one can use a GPD approach as the one presented below to model the severities.

[2]The business lines are corporate finance, trading & sales, retail banking, commercial banking, payment and settlement, agency services, asset management and retail brokerage. The event types are internal fraud, external fraud, employment practices & workplace safety, clients, products & business practices, damage to physical assets, business disruption & system failures and execution, delivery & process management.

3 The theoretical framework

The Loss Distribution Function (LDF) G (whose density is denoted g) is a mixture of two distributions, the frequency distribution p and the severity distribution F (whose density is f):

$$G = \sum_{\gamma=0}^{\infty} p(\gamma) F^{*\gamma}(x;\theta),\ x > 0,\ G = 0,\ x = 0, \tag{3.1}$$

where $*\gamma$ is the γ-order operator of convolution, θ a vector of parameters. The density is given by,

$$g = \sum_{\gamma=0}^{\infty} p(\gamma) f^{\otimes\gamma}(x;\theta),\ x > 0. \tag{3.2}$$

The severity distribution, on which we focus in the current paper, is defined as a mixture of a lognormal distribution on the *corpus*, and a GPD on the right tail (*cauda*). The GPD density is:

$$f^{GPD}(x;u,\beta,\xi)) = \begin{cases} \frac{1}{\beta}\Big(1+\xi\frac{x-u}{\beta}\Big)^{-1-(\frac{1}{\xi})},\ \text{if } x \geq u,\ 1+\xi(\frac{(x-u)}{\beta}) > 0,\ \beta > 0 \\ \frac{1}{\beta}\Big(1-\frac{x-u}{\beta}\Big),\ \text{if } x \geq u,\ \xi = 0 \end{cases}, \tag{3.3}$$

where $u \in \mathbb{R}^{*+}$ is the threshold, $\beta \in \mathbb{R}^*$ is the scale parameter and $\xi \in \mathbb{R}$ the shape parameter.

Therefore, the severity density in our model is as follows:

$$f(x;u,\beta,\xi,\mu,\sigma) = \begin{cases} f^{(corpus)}(x;\mu,\sigma),\ \text{if } x < u \\ f^{(cauda)} = \frac{1}{1-\int_0^u f^{(corpus)}(x;\mu,\sigma)dx} \times f^{GPD}(x;u,\beta,\xi)),\ \text{if } x \geq u \end{cases}, \tag{3.4}$$

where, μ and σ are the lognormal distribution parameters, and f^{GPD} is the density introduced in (3.3).

Remark 3.1. *Modeling both the losses below and above the threshold has two major advantages:*

- *It enables specifying the threshold regarding a model, i.e. the quantile obtained empirically is different from the quantile we have theoretically.*
- *It permits creating a whole distribution we will be able to use in a multivariate approach (Guégan and Hassani (2010)).*

In order to estimate the parameters of the distribution (3.4), we need first to estimate the threshold u. To do so, we carry out a bootstrap method (Hall (1990), Danielsson et al. (2001))

for which we give practical solutions in the next Section. Once u has been found, ξ and β have to be estimated, therefore we implement a method based on the Anderson-Darling statistics (Luceno (2006)). This method maximizes the following relationship to estimate the parameters:

$$A_n^2 = n \int_{-\infty}^{\infty} \frac{(F^{(GPD)}(x;u,\beta,\xi) - S_n(x))^2}{F(x;u,\beta,\xi)(1 - F^{(GPD)}(x;u,\beta,\xi)))} dF^{(GPD)}(x;u,\beta,\xi), \tag{3.5}$$

where, $F^{(GPD)}(x;u,\beta,\xi)$ is the cumulative distribution function of a GPD whose density is given by (3.3), and $n \in \mathrm{R}^+$ the number of observations. In the fifth Section, we compare this method with other traditional estimators and we show that this approach provides interesting results. Then, as we assumed a lognormal distribution to model the central part of the severity distribution (3.4), we implement an Expectation-Maximization algorithm (Dempster et al. (1977), McLachlan and Krishnan (1997)) to estimate the parameters of the truncated distribution. Finally, the λ parameter of the Poisson distribution $p(\gamma;\lambda)$ is estimated using maximum likelihood method. As soon as we have estimated the whole set of parameters, we can build the loss distribution function (3.1) using a convolution method based on the modified Monte Carlo Algorithm given in Appendix B (Fishman (1996), Cruz (2002), Peters et al. (2007) and Shevchenko (2010)), and compute a VaR (Riskmetrics (1993), Artzner et al. (1999)),

$$VaR_{(1-\alpha)\%} = \inf(x \in \mathbb{R} : P(X > x) \leq (1 - \alpha)), \tag{3.6}$$

where, X is a random variable characterized by the previous LDF, and $\alpha = 0.1$.

The Figure 1 illustrates the method introduced in this section through a simulation experiment. We exhibit the histogram of the Historical LDF. The black line corresponds to a LDF mixing a Poisson distribution and a lognormal severity. The associated 99.9% VaR is pointed out by Δ. The dashed line represents a LDF mixing a Poisson distribution and a multiple pattern severity distribution (lognormal-GPD given by (3.4)) using the Monte Carlo algorithm described in Appendix B. The corresponding VaR is represented by a $+$. The right graph focuses on the right tail of the LDFs described in this paragraph. This figure highlights the fact that this method enables thickening up the right tail of the LDF.

Remark 3.2. *Drawing a parallel between this two-behavior approach with other methods such as using g-and-h distributions (Peters and Sisson (2006), Dutta and J. (2007) and Degan et al.*

(2007)), we conclude that it does not exist a method which supplant the others. Trying the g-and-h distribution to model some of our data sets, we had most of the time bad fitting results or unrealistic capital charges, nevertheless in some particular cases the flexibility offered by this distribution could be a reliable solution to model multimodal severity distributions. The method presented in this paper, especially when we have enough data above the threshold and as soon as the goodness-of-fit test validates the distribution, may be a powerful alternative.

4 Discussion around the threshold "u" of the GPD

In this section, we focus on the way to determine an efficient threshold u consistent with the specific data set we work with, that is to say operational risks losses. In order to solve this problem, we assume that we observe a data set $X = (X_1, ..., X_n)$, and we denote $\underline{X} = (X_{(1)}, ..., X_{(n)})$ its order statistics, We first estimate the parameter ξ of the GPD (3.3) with the whole sample using the traditional Hill estimator (Hill (1975)):

$$\hat{\xi}_n(k) = \frac{1}{k}\sum_{i=1}^{k} \log X_{(n-i+1)} - \log X_{(n-k)}, \tag{4.1}$$

k being the number of losses above the threshold. Thus, by the way we introduce a direct relationship between k and u. Indeed, once we have k we have the threshold and *vice versa*. We first determine k using the Hill estimate (4.1) and then implement a bootstrap method to obtain u. The different steps are the following:

1. We plot $\hat{\xi}_n(k)$ with respect to k, as in Figure 2.

2. As soon as we detect a change in the plot pattern ("break"[3]) as in Figure 2 we build an information subset X_1 (size n_1) containing this break and the values above it. For example, in Figure 2 we heuristically determine a break at the value 100 € which corresponds to the initial value of the subsample. Using the values in X_1 we determine the parameter k_1 minimizing the following criterion:

$$AMSE(k_1) := E(\hat{\xi}_{n_1}(k_1) - \xi)^2 \tag{4.2}$$

[3]The break is determined empirically by finding a change in the slop.

Documents de Travail du Centre d'Economie de la Sorbonne - 2010.96 (Version révisée)

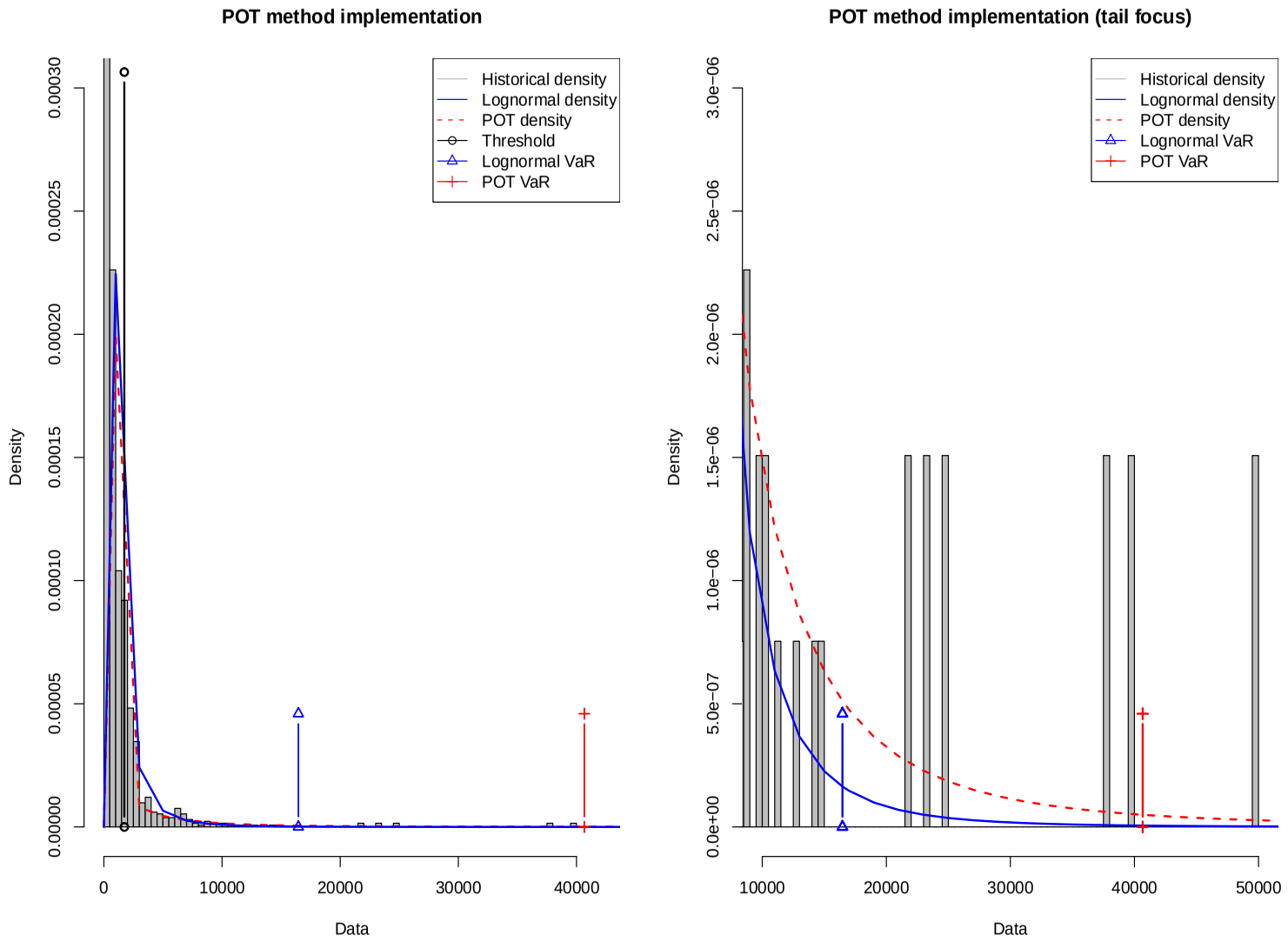

Figure 1: Method Illustration: This figure presents the histogram (in grey) of the Historical LDF. The black line corresponds to a LDF mixing a Poisson distribution and a lognormal severity. The associated 99.9% VaR is pointed out by Δ. The dashed line represents a LDF mixing a Poisson distribution and a multiple pattern severity (lognormal-GPD given by (3.4)) using the algorithm discribed in Appendix B. The corresponding VaR is represented by a $+$. The right graph focuses on the right tail of the LDFs described above. This figure highlights the fact that this method enables thickening up the right tail of the LDF.

3. In the relation (4.2), the parameter ξ is unknown and we estimate it bootstrapping $\hat{\xi}_{n_1}(k_1)$ as follows. From the set previous X_1, we draw J ($J \in \mathbb{N}$) subsamples with replacement of size n_2 ($n_2 < n_1$), $(X_1^{(j)}, ..., X_{n_2}^{(j)})$, $j = 1, ..., J$. Their corresponding order statistics are $(X_{(1)}^{(j)}, ..., X_{(n_2)}^{(j)})$ and the corresponding bootstrapped Hill estimate of $\hat{\xi}_{n_1}(k_1)$ from X_1 is equal to,

$$\hat{\xi}_{n_2}(k_2) = \frac{1}{k_2} \sum_{i=1}^{k_2} \log \underline{X}_{(n_2-i+1)} - \log \underline{X}_{(n_2-k_2)},\ k_2 = 1, ..., n_2, \tag{4.3}$$

where

$$\underline{X}_l = \frac{1}{J} \sum_{j=1}^{J} X_{(l)}^{(j)},\ l = 1, ..., n_2. \tag{4.4}$$

4. Now, replacing ξ by $\hat{\xi}_{n_2}(k_2)$ in (4.2), we minimize the following criterion with respect to k_2.

$$\widehat{AMSE}(k_2) := E(\hat{\xi}_{n_1}(k_1) - \hat{\xi}_{n_2}(k_2))^2. \tag{4.5}$$

5. Then, given (k_1, k_2), and following Hall (1982; 1990) we estimate[4] the threshold u by,

$$u = \left\lceil k_2 * \left(\frac{n_1}{n_2}\right)^{\frac{2}{3}} \right\rceil. \tag{4.6}$$

In order to implement the previous methodology, we need to pay attention to the following points:

1. If an obvious pattern change (from erratic to linear for example) is observed in the representation of $\hat{\xi}_n(k)$ with respect to k (Step 1), we can directly read u on the graph.

2. In the step 4, k_1 is fixed. In practice, we can hesitate between several k_1-values, denoted $\left(k_1^1, ..., k_1^\Lambda\right)$, $\Lambda \in \mathbb{N}$, which means that for each $k_1^{(j)}$, $j = 1, ..., \Lambda$, we obtain a k_2-value denoted $k_2^{(j)}$. For each couple $(k_1^{(j)}, k_2^{(j)})$, we choose the $k_2^{(j)}$-value, for which the relationship (4.5) is the "'most stable"', and therefore, we compute the criteria (4.5) for d successive values of k_2 taken around $k_2^{(j)}$. In that case the parameter k_2 is obtained minimizing:

$$\phi(k_2^{(j)}, d) = \sqrt{\sum_{i=1}^{d} \left(M_1^{(j)} - M_i^{*(j)}\right)^2},\ j = 1, ..., \Lambda, \tag{4.7}$$

where, $M_1^{(j)} \equiv \widehat{AMSE}(k_2^{(j)})$ is obtained from (4.5) for each $k_2^{(j)}$, and $M_i^{*(j)}$ is computed using the i value around $k_2^{(j)}$, $i = 1, ..., d$.

[4]We are only working on a part of the initial sample, it is necessary to adjust k_2 to the whole data set.

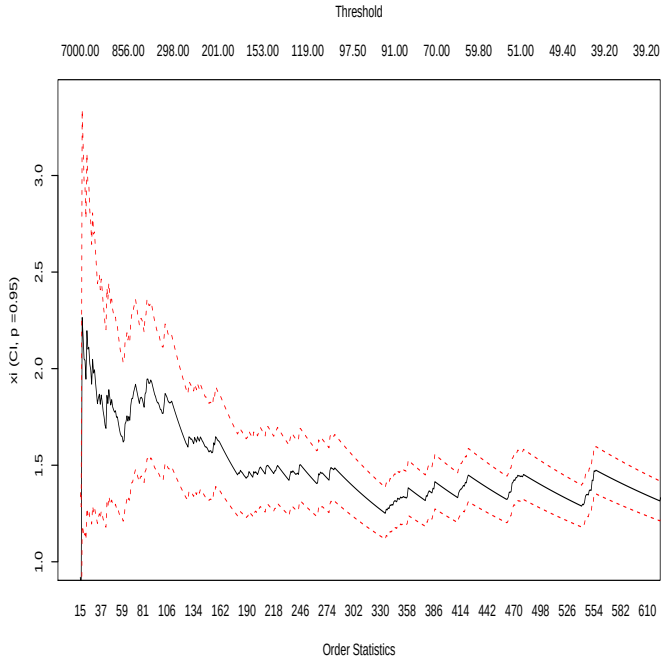

Figure 2: Using the 610 highest values of the Basel case (Retail Banking; Damage to Physical Assets) of 2007, we represent the Hill estimates considering thresholds varying between 39.2 and 7000 €. The graph becomes roughly stable around the value 180, but even if we are able to identify the zone in which the threshold would stand, we have not yet an exact value, thus to estimate u, we define the set X_1 from the value 100 €, thus X_1 contains 290 data.(Upper and lower dashed lines represents a 95% confidence interval.)

3. If we cannot interpret the Hill plot (Step 1), either because the data set is too unstable or too stable, we cannot use the previous algorithm, therefore, we use the whole information set X, and follow the method developed in Danielsson et al. (2001) to estimate the ξ parameter (this method does not require an *a priori* value of the threshold). We estimate ξ by:

$$\tilde{\xi}_n(k) = \frac{\frac{1}{k}\sum_{i=1}^{k}(\log X_{(n-i+1)} - \log X_{(n-k)})^2}{2\hat{\xi}_n(k)}. \tag{4.8}$$

For a certain[5] n_2, we bootstrap $\tilde{\xi}_n(k)$ which appears in relationship (4.8) following the step 3 of the previous algorithm to create $\tilde{\xi}_{n_2}(k_2)$, where the parameter k_2 is obtained

[5]A short algorithm providing the optimal n_2 is given in Danielsson et al. (2001).

minimizing,

$$\widetilde{AMSE}(k_2) := E(\hat{\xi}_{n_2}(k_2) - \tilde{\xi}_{n_2}(k_2))^2. \tag{4.9}$$

Then for a given k_2, u is equal to:

$$u = \frac{(k_2)^2}{k_3}\left(\frac{(\log k_2)^2}{(2\log n_2 - \log k_2)^2}\right)^{\frac{\log n_2 - \log k_2}{\log n_2}}, \tag{4.10}$$

where k_3 is obtained exactly as k_2 with $n_3 = \frac{n_2^2}{n}$.

4. Shevchenko and Temnov (2009) suggest the use of a time varying threshold. In the current paper, we use a static methodology, nevertheless it would be interesting to integrate dynamics in the process of estimation. To do so, a possibility would be to use the dynamics of the severity data set given for instance by the incidents' amounts and their occurrence dates. In that case, the results would take into account the behavior changes in the risk faced through time and adjust with the state of nature they are currently in (crises, boom etc.).[6]

5 Capital Requirement

We applied the previous methodology to real operational risks data. We use two data sets denoted $X^{(1)}$ and $X^{(2)}$.

- The set $X^{(1)}$ represents the severity of the business line "Retail Banking" and the event type "Business Disruption & System Failure and Execution" for the year 2007.
- The set $X^{(2)}$ represents the severity of the business line "Payment & Settlement" and the event type "Delivery, Execution and Process Management" for the year 2006.

Their statistics are given in Table 1. We deduce that their corresponding empirical severity distributions are asymmetric (right skewed) and leptokurtic. This statement strengthens our decision to implement a method which cares about extreme values in the distributions.

[6]Nevertheless in the current paper, this methodology cannot be considered because we do not have the corresponding information set. Besides, in Guégan and Hassani (2011) we dealt with a dynamic threshold using a extreme distribution (Fisher and Tippett (1928)) with another data set.

Moments	$X_{(1)}$	$X_{(2)}$
Mean	958.4922	1265.871
Standard Deviation	5010.661	10852.48
Skewness	29.84618	10.62595
Kurtosis	1163.478	119.5531

Table 1: First four moments of the two data sets $X^{(1)}$ and $X^{(2)}$. The third order empirical moment is non null and the fourth order empirical moment is far from 3.

In order to calibrate the appropriate loss distribution function on each data set, we use a Poisson distribution to model the frequency data. In the current paper, we did not challenge the choice of a Poisson distribution to model the frequency as the gain obtained from a continuous approximation or the use of another theoretical distribution has almost negligible impact on capital computation. Therefore, estimating the λ parameters of the Poisson distribution by maximum likelihood method, we get $\lambda^{(1)} = 2653$ and $\lambda^{(2)} = 1292$ respectively on the sets $X^{(1)}$ and $X^{(2)}$.

To fit the severities, we tried several usual distributions suggested in the literature, but the Kolmogorov-Smirnov test rejected all (Table 6). Therefore, we continue our exercise and focus on the estimation of the threshold. To estimate the threshold u of the GPDs in the relation (3.4), we define two new sets $X_1^{(1)}$ an $X_1^{(2)}$, following the notations introduced in the previous section. Their sizes are respectively $n_1^{(1)} = 531$ and $n_1^{(2)} = 194$. When we bootstrap the data, we used respectively two sets of size $n_2^{(1)} = 266$ and $n_2^{(2)} = 97$. The resulting thresholds are $u^{(1)} = 1645.07$ and $u^{(2)} = 179$. They imply respectively 191 and 315 data above these thresholds. We denote $X^{(GPD_1)}$ and $X^{(GPD_2)}$ the two corresponding data sets on which we fit the GPDs.

We estimate the parameters ξ and β of the GPDs defined in (3.4) using the method introduced by (Luceno (2006)) given in (3.5), denoting this method M1. We also consider three other alternative estimation methods to estimate these parameters in order to check their impact on VaR computations. These ones are respectively the Pickands method (M2) (Pickands (1975)), the Hill method (M3) (Hill (1975)), and the Maximum Likelihood method (M4).

We provide in Table 2 the estimations of GPDs parameters (with their standard deviation in brackets) obtained for both sets, using these four methods. Then we estimate the parameters of the lognormal distributions for the sets $X^{(1)} - X^{(GPD_1)}$ and $X^{(2)} - X^{(GPD_2)}$ and the results are provided in Table 3.

Method	$X^{(1)}$		$X^{(2)}$	
	β	ξ	β	ξ
M1	136.678	2.085	932.854	0.767
	(36.91)	(0.687)	(83.71)	(0.101)
M2	133.092	2.173	682.615	1.144
	(39.70)	(0.361)	(160.70)	(0.266)
M3	171.5	1.581	1007	0.66
	(26.19)	(0.668)	(214.36)	(0.228)
M4	159.535	1.948	904.087	0.827
	(27.3755)	(0.21)	(92.31)	(0.097)

Table 2: Estimations of the GPD's parameters ξ and β given in (3.3), using four methods respectively for $u^{(1)} = 1645.07$ and $u^{(2)} = 179$. We provide in brackets the standard deviations computed by bootstrapping.

$X^{(1)} - X^{(GPD_1)}$		$X^{(2)} - X^{(GPD_2)}$	
$\mu = 5.702144$	$\sigma = 1.103373$	$\mu = 3.593098$	$\sigma = 1.510882$
(0.07134)	(0.03016)	(0.04112)	(0.02576)

Table 3: Estimations of the parameters of the lognormal distributions fitted on the data sets $X^{(1)} - X^{(GPD_1)}$ and $X^{(2)} - X^{(GPD_2)}$. Standard deviation are provided in brackets.

As soon as all the parameters have been estimated, we implement a Monte Carlo algorithm with a million of iterations to build G, and we use it to compute a 99.9% VaR which provides the required amount of capital. Table 4 provides the VaRs of the set $X^{(2)}$ using the four estimation methods to estimate the threshold. We observe miscellanous amounts, varying from 5 725 341 euros to 538 480 990 euros. This last result is obtained from a distribution whose mean

is infinite (the parameter ξ is greater than 1). This result points that the choice of the distribution and the estimation procedure are crucial for the VaR computations. The M1 method has the advantage to maximize the goodness-of-fit of the GPD and insures that the distribution cannot be statistically rejected. Therefore, one can argue that the VaR at 15 700 112 euros could be a reasonnable value considering the information set we have, and the assumptions done.

We do not provide the VaRs for the set $X^{(1)}$ as the results we obtained imply a simulated VaR multiplied by more than one hundred compared to the historical one. Here again, we can observe that the underlying model has an infinite mean.

Analyzing the results and controlling every step of the method, we concluded that the difficulty to obtain a "correct" and a "robust" model comes from the structure of the data set. Considering that different kinds of data are mixed in a single set it appears difficult to fit an accurate distribution on the severities. In order to analyze the impact of the granularity strategy, it would be interesting to split the data in multiple subsets whether the information sets sizes are large enough. In that latter case the computed VaRs would reflect risks created by specific kinds of risk, and finally aggregating them we could compare the results to the largest level of granularity.[7]

Method	VaR
M1	15 700 112 euros
M2	538 480 990 euros
M3	5 725 341 euros
M4	27 944 558 euros

Table 4: VaR estimations for $u = 179$, $\mu = 5.681191$ and $\sigma = 1.081609$ for the data set $X^{(2)}$. Estimations have a tremendous impact on the VaRs. In this table, these ones may differ by 9405%.

Finally we want to notice the influence of the threshold on the estimation of the GPD's parameters β and ξ. This effect is due to the fact that the information set changed. We illustrate this

[7] We do not provide those results as we did not have the corresponding data sets.

point providing estimates for five threshold values, for the set $X^{(1)}$ (Table 5). All the estimated values are relevant despite a great instability. Thus the corresponding VaR and capital amount would be different. All steps of the previous methodology: estimation of u, estimation of the parameters ξ and β, methods of estimation play an important role in the computation of the capital requirement.

Threshold u	β	Standard error β	ξ	Standard error ξ
1162.12	1110.8538	87.77	0.6026	0.06877
1608.27	946.6787	94.68	0.7924	0.09329
1645.07	904.087	92.31	0.827	0.09713
2021.48	1188.2956	155.48	0.9149	0.1397
2177.58	1385.5295	199.56	0.8879	0.14

Table 5: This table presents GPD's parameters estimates and their standard errors for different thresholds u for the data set $X^{(2)}$. Parameters are very sensitive to the threshold value.

Conclusion

In this paper, we focus on the estimation of the LDF used to estimate operational risks. We highlight that the VaR computation depends on the method implemented. Therefore, a sharp analysis of the LDF is essential, and even if the LDF is a convolution of two distinct distributions, the capital requirements seem particularly sensitive to the choice and the estimation of the severity distribution. We present a new method to compute the severity distribution, splitting it in two parts in order to better take into account the large losses. We use a GPD on the right tail for which we provide innovative theoretical and practical solutions, and fit a lognormal distribution on the remaining data. Then, to build the final LDF, we apply a new adapted Monte Carlo algorithm. This method enables capturing at least two distinct behaviors in a severity distribution, and sometimes a bit more if the multimodality is not too strongly marked. We point out the importance of the Basel matrix construction and the so-called granularity effect.

Second, we underline the fact that once the threshold of the GPD has been found, the method chosen to estimate GPD's parameters tremendously impacts the VaR: it seems that this fact has never been discussed before.

In our applications, we have been confronted to an important problem arising from the data bases construction. Following the regulator, who demands to take into account the right tail of the severity distribution to estimate the LDF, we sometimes obtained unrealistic VaRs (i.e. equal to half the bank capitalization) due to infinite mean models. On the other hand, if we do not care about the severity tail, we have negligible VaRs. Then, even if it seems reasonable to follow regulator requirements, we point out two important facts: either the Bank are overexposed to these risks or the data sets are badly built. We privilege the last point and ask operational risks managers to pay attention to the large diversity of data origins. The correct question is to propose another way for building the data sets used to compute operational risks taking into account the difference which exists inside the different risk categories. It seems that it has no sense to mix some risks together. In our knowledge, this problem has not yet been discussed and the regulators need to reexamine risk typology and data sets granularity on which are computed operational risk capital requirements.

The method we provide enables to compute capital requirements with respect to Basel accords, so, we strictly stick to them. In this paper we focused on the effectiveness of the proposed method on an operational standpoint, however, other research tracks can extend the results developed in this paper. Even if in this paper we focus on the VaR computation to obtain the corresponding capital requirement, it will be interesting to use other risks measures, for instance the Expected Shortfall (ES) measure, and analyze their impact on the computation of the associated capital requirement. The use of the ES is not required by the regulator but it would be interesting because it is a coherent risk measure (Artzner et al. (1999), Extensions are proposed in Guégan and Hassani (2011)). On the other hand there are alternatives to the Monte Carlo method to calculate the LDF (Luo and Shevchenko (2009), Guégan and Hassani (2009)) permitting to improve the computation speed that could be investigated. Finally, to take into account the effect of the tail behavior, we focus in this paper on the Generalized Pareto Distribution, nevertheless other alternatives could be considered, e.g. the extreme value distributions defined in the Fisher-Tippett theorem (Fisher and Tippett (1928)), the g-and-h distributions, or multivariate methods through the extreme value copulas (Gudendorf and Segers (2010)). All these subjects will be discussed in companion papers.

References

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

BIS (2001), 'Working paper on the regulatory treatment of operational risk', *Bank for International Settlements* .

BIS (2004), 'International convergence of capital measurement and capital standards', *Bank for International Settlements* .

Chernobai, A., Rachev, S. T. and Fabozzi, F. J. (2007), *Operational Risk: A Guide to Basel II Capital Requirements, Models, and Analysis*, John Wiley & Sons.

Cruz, M. (2002), *Modeling, measuring and hedging operational risk.*, Wiley, London.

Cruz, M. (2004), *Operational Risk Modelling and Analysis*, Risk Books, London.

Danielsson, J., de Haan, L., Peng, L. and de Vries, C. (2001), 'Using a bootstrap method to choose the sample fraction in tail index estimation', *Journal of Multivariate Analysis* **76**, 226–248.

Davis, R. and Resnick, S. (1988), 'Tail estimates motivated by extreme value theory', *Annals of Stat.* **12**, 1467 – 1468.

Degan, M., Embrechts, P. and Lambrigger, D. (2007), 'The quantitative modeling of operational risk: between g-and-h and evt.', *Astin Bulletin* **37(2)**, 265–291.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977), 'Maximum likelihood from incomplete data via the em algorithm', *Journal of the Royal Statistical Society: Series B* **39**, 1 – 38.

Dutta, K. and J., P. (2007), 'A tale of tails: an empirical analysis of loss distribution models for estimating operational risk capital.', *Working Paper N°. 06-13, Federal Reserve Bank of Boston* .

Embrechts, P., Klüppelberg, C. and Mikosh, T. (1997), *Modelling Extremal Events: for Insurance and Finance*, Springer, Berlin.

Fisher, R. A. and Tippett, L. H. C. (1928), 'Limiting forms of frequency distributions of the largest or smallest member of a sample', *Proc. Cambridge Philosophical Soc.* **24**, 180 – 190.

Fishman, G. S. (1996), *Monte Carlo*, Springer-Verlag, Berlin.

Frachot, A., Georges, P. and Roncalli, T. (2001), 'Loss distribution approach for operational risk', *Working Paper, GRO, Crédit Lyonnais, Paris* .

Gudendorf, G. and Segers, J. (2010), *Extreme-value copulas. In: Proceedings of the Workshop on Copula Theory and Its Applications held in Warsaw (editor: P. Jaworski)*, Springer Media, New York.

Guégan, D. and Hassani, B. (2010), 'n-dimensional copula contributions to multivariate operational risk capital computations.', *Working Paper, University Paris 1* **2**.

Guégan, D. and Hassani, B. (2011), 'A mathematical resurgence of risk management: an extreme modeling of expert opinions.', *Working Paper, University Paris 1* .

Guégan, D. and Hassani, B. K. (2009), 'A modified panjer algorithm for operational risk capital computation', *The Journal of Operational Risk* **4**, 53 – 72.

Hall, P. (1982), 'On some simple estimates of an exponent of regular variation', *J.R.S.S.B.* **44**, 37 – 42.

Hall, P. (1990), 'Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems.', *Journal of multivariate analysis* **32**, 177–203.

Hill, B. M. (1975), 'A simple general approach to inference about the tail of a distribution', *Ann. Statist.* **3**, 1163 – 1174.

Luceno, A. (2006), 'Fitting the generalized pareto distribution to data using maximum goodness-of-fit estimators', *Computational statistics and data analysis* **51**, 904 – 917.

Luo, X. and Shevchenko, P. (2009), 'Computing tails of compound distributions using direct numerical integration', *Journal of Computational Finance* **13**, 73–111.

McLachlan, G. and Krishnan, T. (1997), *The EM algorithm and extensions*, John Wiley & Sons, London.

Peters, G. and Sisson, S. (2006), 'Bayesian inference, monte carlo sampling and operational risk.', *Journal of Operational Risk.* **1(3)**, 27–50.

Peters, G. W., Johansen, A. M. and Doucet, A. (2007), 'Simulation of the annual loss distribution in operational risk via panjer recursions and volterra integral equations for value at risk and expected shortfall estimation.', *Journal of Operational Risk* **2**, 29–58.

Pickands, J. (1975), 'Statistical inference using extreme order statistics', *annals of Statistics* **3**, 119–131.

Riskmetrics (1993), 'Var', *JP Morgan* .

Shevchenko, P. (2010), 'Calculation of aggregate loss distribution approach for operational risk.', *The Journal of Operational Risk* **5(2)**, 3–40.

Shevchenko, P. and Temnov, G. (2009), 'Modeling operational risk data reported above a time-varying threshold.', *The Journal of Operational Risk* **4(2)**, 19–42.

Shevchenko, P. V. (2011), *Modelling Operational Risk Using Bayesian Inference*, Springer, Berlin.

A Traditional distributions

$X^{(1)}$			$X^{(2)}$		
Distribution	Statistics	P-value	Distribution	Statistics	P-value
lognormal ($\mu = 5.86$, $\sigma = 1.25$)	0.1498	$<$ 2.2e-16	lognormal ($\mu = 3.69$,$\sigma = 1.78$)	0.1066	3.552e-13
exponential ($\eta = 1.04e-03$)	0.306	$<$ 2.2e-16	exponential ($\eta = 7.9e-04$)	0.731	$<$ 2.2e-16
Weibull ($\beta = 652.79$, $\xi = 0.704$)	0.2035	$<$ 2.2e-16	Weibull ($\beta = 91.82$, $\xi = 0.41$)	0.2238	$<$ 2.2e-16
Gumbel ($\nu = 392.46$, $\xi = 680.01$)	0.2833	$<$ 2.2e-16	Gumbel ($\nu = 116.16$, $\xi = 1191.18$)	0.4904	$<$ 2.2e-16
Fréchet ($\nu = 238.58$, $\beta = 222.87$ $\xi = 0.67$)	0.1267	$<$ 2.2e-16	Fréchet ($\nu = 21.73$, $\beta = 28.49$ $\xi = 1.14$)	0.0929	4.072e-10

Table 6: Goodness-of-fit of the presented theoretical distributions to the sets $X^{(1)}$ and $X^{(2)}$: the parameters have been estimated by maximum likelihood and we have chosen the Kolmogorov-Smirnov statistic. The p-values are all lower than 1%. It's not satisfactory. Considering goodness-of-fit test results, it is not possible to state that a distribution is better than another to model the losses.

B An adapted Monte Carlo algorithm

1. We simulate n realizations of the frequency distribution, $q_1, ..., q_n$, $n \geq 1000000$.

2. For each q_i, $i = 1, \cdots, n$, we simulate q_i values of the central distribution, $(s_1, ..., s_{q_i})$, and compare them to the threshold:

   (a) If $s_j < u$, $\forall j \in [1, q_i]$ we keep them.

   (b) If $s_j \geq u$, we count the number of exceedences, $(z_1, ..., z_n)$ for each frequency q_i and for each i, we draw z_i realizations of the GPD, $(s_1^{(GPD)}, ..., s_{z_i}^{(GPD)})$.

3. Finally, we have $G_i = \sum_{j=1}^{q_i}(s_j + s_j^{(GPD)})$, a realization of the LDF, $i = 1, ..., n$.

4. The set $(G_1, ..., G_n)$ provides an empirical approximation of the LDF.