Kernel Alternatives to Approximate Operational Severity Distribution: An Empirical Application.

Filippo di Pietro
University of Seville
fdi@us.es
María Dolores Oliver Alfonso
University of Seville
moliver@us.es
Ana I. Irimia Diéguez
University of Seville
anairimia@us.es

The estimation of severity loss distribution is one the main topic in operational risk estimation. Numerous parametric estimations have been suggested although very few work for both high frequency small losses and low frequency big losses. In this paper several estimation are explored. The good performance of the double transformation kernel estimation in the context of operational risk severity is worthy of a special mention. This method is based on the work of Bolancé and Guillén (2009), it was initially proposed in the context of the cost of claims insurance, and it means an advance in operational risk research.

Keys words: Operational risk, Basel II, Loss Distribution Approach, Non-parametric estimation, Kernel smoothing estimation.
JEL Classification: G10, G20, G21, G32, C11, C14, C16

1. Introduction

The revised Basel Capital Accord requires banks to meet a capital requirement for operational risk as part of an overall risk-based capital framework. With regards to the definition aspects, the Risk Management Group (RMG) of the Basel Committee and industry representatives have agreed on a standardized definition of operational risk, i.e., “the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events” (BIS, 2001). The discipline proposed establishes various schemes for the calculation of the operational risk charge, which increases sophistication and risk sensitivity. The most sophisticated approach is the Advanced Measurement Approaches (AMA), based on the adoption of the internal models of banks.

Concerning the measurement issue, a growing number of articles, research papers and books have addressed the topic from a theoretical point of view. In practice, this objective is complicated by the relatively short period over which operational risk data have been gathered by banks. Obviously, the greatest difficulty is in collecting information on infrequent, but large losses, which, on the other hand, contribute the most to the capital requirement. The need to evaluate the exposure to potentially severe tail events is one of the reasons why the new Capital framework requires banks to supplement internal data with further sources, (i.e., external data, scenario analysis), in order to compute their operational risk capital requirement.

Recently, the measurement of operational risk has moved towards the data-driven Loss Distribution Approach (LDA) and therefore, many financial institutions have begun collecting operational loss data in order to take advantage of this approach. The LDA approach requires the aggregation of the severity and frequency distributions in order to obtain the aggregated loss distribution.

The estimation of the severity loss distribution is probably one of the most significant phases and that which involves the highest number of complications towards the estimation of the capital requirement of operational risk. An incorrect estimation of distribution leads to a severe distortion of the model and a low estimate or overestimates of regulatory capital with respect to the operational risk. This could have a big impact on the Basel II economic capital requirement. Recent literature on operational risk has focused attention on the use of a parametric estimation of loss distribution. This is the simplest method to follow, since it attempts to fit analytical distributions with certain properties. The aim of this approach is to find a distribution of losses that may be feasible to the severity distribution of the losses of the sample available. Another commonly applied technique in operational risk is the Extreme Value Theory (EVT) which is a good methodology in cases where the main attention is the tail of the distribution.

We take as alternative non-parametric estimation which permit the quantification of operational loss severity by fitting the whole distribution and that do not require the specification of parametric estimation. With this in mind, the kernel estimation is taken as the starting point which has been improved with the parametric transformation approach given in Wand, Marron and Ruppert (1991) and recently considered in Bolancé, Guillén and Nielsen, (2008), Buch-Larsen, Guillén, Nielsen, and Bolancé (2005). The analysis is completed by using the latest methodology developed by Bolancé and Guillén (2009), based on a double transformation which in our opinion can notably improve the operational loss severity estimation. In order to explore all possible methods, a data sample of losses is utilized based on the operational risk of a medium-sized Spanish Savings Bank.

In this article, we attempt to find which methodology of estimation (parametric and non-parametric) yields the most appropriate measure of operational risk severity, with a particular focus on the case of Savings Bank. We demonstrate that non-parametric estimation with this improvement (the double transformation kernel estimation) is a good alternative methodology for the approximation of the loss severity distribution, since it performs much better than parametric estimation. These methodologies render the estimation of a threshold unnecessary. Another good property of the non-parametric estimation with respect to EVT is that it seems to not overestimate the capital requirement. The double transformation kernel estimation was applied in the insurance claim field to approximate loss distribution. We believe that this methodology can also represent an advance in operational loss severity estimation.

The remainder of the paper is organized in the following way. In the second section, a description of the parametric and non parametric methodologies is reported. In the third section, the characteristics of data and an exploratory analysis are presented. In the fourth section, a comparison of the distributions obtained with parametric and non-parametric estimations to approximate the severity loss distribution is included. In the fifth section, an operational VaR is estimated aggregating the different severity distributions with a Poisson for the frequency distribution. The conclusion is the last section.

2. Theoretical Framework

In operational risk and specifically in the dataset considered, most operational losses are small and extreme losses are rarely observed, although the latter can have a substantial influence on the capital charge estimation.

In previous studies, i.e., de Fontnouvelle et. al., (2003), Frachot et al. (2003), Dutta and Perry (2006), the authors have tried to find parametric methods that could fit the whole distribution well. However, it is a major problem to find a distribution that provides a good fit for both the body and the tail part of the loss distribution. This method suffers from another problem, since no emphasis is given to the importance of a good fit in the tail, where the problem of operational risk is focused.

One way to solve such an inconvenience is to analyze small and large losses separately with the Extreme Value Theory (EVT). This approach involves some drawbacks: choosing the appropriate parametric model; finding the best way to estimate the parameter; and, most importantly, identifying a criterion to determine the threshold not to mention that, in this case, any loss threshold applied would lose important information and, therefore, the robustness of the results obtained.

A non-parametric approximation, especially by kernel estimation, can be a good alternative. The classic kernel estimation of the distribution function is a simple smoothing of the empirical distribution function, and for this reason a lack of sample information implies a loss of precision in the approximation of the heavy losses. The transformation kernel estimation is an alternative that improves the results obtained with the classic kernel estimation. This approach transforms the variable by using a concave function symmetrization of the original data, and then obtains the classic transformed kernel variable.

The methodology used in Bolancé and Guillén (2009) is proposed as an alternative method. They propose a new transformation kernel estimation, based on a double transformation, which can improve the estimation of risk measures. In this paper this methodology is tested in the field of operational risk in order to compare the results with those obtained through more traditional methodologies. With this in mind, in addition to the non-parametric methodology to fit the data, parametric methodologies by fitting distributions such as lognormal, Weibull and the generalized Pareto are tested.

2.1. The parametric loss models.

This section explores three parametric alternatives which capture the severity losses for operational risk.

The most common distribution in modelling the OR is the lognormal whereby distribution[1]. One can say that a random variable X has a lognormal distribution if its density and distribution are, respectively[2]:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma x}} e^{-\frac{(\log x - u)^2}{2\sigma^2}} \quad (1)$$

$$F(x) = \Phi\left(\frac{\log x - u}{\sigma}\right), \ x>0 \quad (2)$$

where Φ(x) is the standard normal distribution.

The parameters $u(-\infty < u < \infty)$ and $\sigma(\sigma > 0)$ are the centre and scale, respectively, and can be estimated with the maximum likelihood method:

$$\hat{u} = \frac{1}{n}\sum_{j=1}^{n} \log x_j \ , \hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n} (\log x_j - \hat{u})^2 \quad (3)$$

The inverse distribution is $F^{-1}(p) = e^{\Phi^{-1}(p)\sigma + u}$, therefore the lognormal random variable can be simulated.

The Weibull distribution is a generalization of the exponential distribution, whereby two parameters instead of one allow more flexibility and heavier tails. The density and distribution are:

[1] The lognormal distribution was proposed by the Basel Committee for modelling operational risk.

[2] See: Chernobai, Rachev, and Fabozzi (2006).

$$f(x) = \alpha\beta x^{\alpha-1} e^{-\beta x^{\alpha}} \tag{4}$$

$$F(x) = 1 - e^{-Bx^{\alpha}} \tag{5}$$

where x> 0, with $\beta(\beta > 0)$ as the scale parameter and $\alpha(\alpha > 0)$ as the shape parameter.

There is no inverse of a Weibull random variable in a closed formula. To generate a Weibull random variable, an exponential random variable Y must be generated with parameter β and then follow the transformation $X = Y^{1/\alpha}$.

2.2. Extreme Value Theory

As seen in the conventional inference, the influence of the small/medium-sized losses in the curve parameters estimation prevents the attainment of models that fit the tail data accurately. An obvious solution to this problem is to disregard the body of the distribution, and to focus the analysis only on the large losses. When only the tail area is considered, several distributions could be adopted, such as lognormal and Pareto, which are often used in insurance to model large claims. However, in this section the attention is focused on extreme distribution stemming from the Extreme Value Theory (EVT) [3], and especially on Peak Over Threshold (POT)[4].

As witnessed by Chapelle et al. (2008), this approach enables us to simultaneously determine the cut-off threshold and to calibrate a distribution for extreme losses above this threshold. The severity component of the POT method is based on a distribution (Generalized Pareto Distribution - GPD), whose cumulative function is usually expressed as the following two-parameter distribution[5]:

$$GPD_{\xi,\sigma}(x) = \begin{cases} 1 - (1 + \xi \dfrac{x}{\sigma})^{-1/\xi} & if \ \xi \neq 0 \\ 1 - \exp\left\{-\dfrac{x}{\sigma}\right\} & if \ \xi = 0 \end{cases} \tag{9}$$

---

3 For a comprehensive source on the application of EVT in finance and insurance see Embrechts *et al.*, (1997), and Reiss and Thomas, (2001).

4 In the *Peak Over Threshold* (POT) model, the focus of the statistical analysis is placed on the observations that lie above a certain high threshold.

5 See: Moscadelli (2004).

where: $x \geq 0 \; if \; \xi \geq 0, 0 \leq x \leq -\sigma/\xi \; if \; \xi < 0$ and ξ and σ represent the shape and the scale parameter respectively.

In this work we use the extended version of GPD which includes a location parameter u :

$$GPD_{\xi,\sigma}(x) = \begin{cases} 1-(1+\xi\dfrac{x-u}{\sigma})^{-1/\xi} & if \; \xi \neq 0 \\ 1-\exp\left\{-\dfrac{x-u}{\sigma}\right\} & if \; \xi = 0 \end{cases} \tag{10}$$

The inverse of the GPD distribution has a simple form:

$$\begin{aligned} &F^{-1}(p) = u - \beta\log(1-p) \text{ for } \xi = 0 \text{ and} \\ &F^{-1}(p) = u - \beta(1-p^{-\xi})/\xi \text{ for } \xi \neq 0 \end{aligned} \tag{11}$$

Although, several authors (see Dupuis, 2004; Matthys and Beirlant, 2003) have suggested methods to identify the threshold, there is no single widely-accepted method to select the appropriate cut-off.

2.3. The Kernel Estimation of severity loss distribution

The estimation of the kernel density function is a non-parametric method to approximate the probability function of a random variable. It is expressed as:

$$\hat{f}_X(x) = \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{x-X_i}{h}\right), \tag{12}$$

Where k(.) is the kernel function or weight function that satisfies certain regularity conditions, and is usually a symmetric density function such as normal distribution, centred at zero and asymptotically bounded or unbounded and $\{h_n\}$ is a positive constant sequence known as window width, smoothing parameter.

In this work the Epanechnikov Kernel is used:

$$k(t) = \begin{cases} \frac{3}{4}(1-t^{2)} & if \; |t| \leq 1 \\ 0 & if |t| > 1 \end{cases} \tag{13}$$

By integrating (12) and using the change of variable $t = (u - X_i)/h$, we obtain the kernel estimation of the distribution function:

$$\hat{F}_X(x) = \int_{-\infty}^{x} \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{u - X_i}{h}\right) du = \frac{1}{nh}\sum_{i=1}^{n}\int_{-\infty}^{x} k\left(\frac{u - X_i}{h}\right) du = \frac{1}{nh}\sum_{i=1}^{n}\int_{-\infty}^{\frac{x-L_i}{h}} k\left(\frac{u - X_i}{h}\right) du \quad (15)$$

The properties of the kernel estimation of the distribution function were analyzed by Reiss (1981) and Azzalini (1981). Both authors suggest that when $n \to \infty$, the mean square error of $\hat{F}_X(x)$ is approximated as:

$$\frac{F_X(x)(1 - F_X(x))}{n} - \frac{hf(x)(1 - \int K^2(t)dt}{n} + h^4\left(\frac{1}{2} f_X'(x)\int t^2 k(t)dt\right)^2 \quad (16)$$

The first two parts of the above expression correspond to the asymptotic variance and the third term is the squared asymptotic bias.

2.4. Transformation Kernel Estimation

Wand, Marron and Rupert (1991) set out in their work, that the classical kernel estimate of the density function is a good alternative for densities that are shaped similarly to a Gaussian distribution function. However, this estimate may have problems when the shape of the estimated densities are far from being Gaussian. They propose, as a solution, transforming the data prior to the estimated core, and then re-transform the data to bring them into the original scale. This procedure substantially improves the performance of the classical kernel estimate, and obtains results similar to the estimate kernel density function with a different bandwidth (smoothing parameter), but estimating only one bandwidth parameter for the whole function. Wand, Marron and Rupert (1991) use "shifted power transformation family" to transform the data. Burch-Larsen et al. (2005) use the distribution function of a generalized Champernowne distribution to transform the data:

$$T_{\alpha,M,c}(x) = \frac{(x+c)^\alpha - c^\alpha}{(x+c)^\alpha + (M+c)^\alpha - 2c^\alpha} \quad \forall x \in R_+ \quad (17)$$

This methodology is used by Gustafsson et al. (2006) in a context of operational risk in insurance companies. In particular, the author intends to find a methodology to model both the body and tail of the distribution of severity.

The transformations proposed in the previously cited works are carried out in a context of density function.

To estimate the distribution function, which is the goal of this work, not many authors use the transformation kernel estimation. Swanepoel (2005) propose a kernel estimator of the distribution function based on a non-parametric transformation of data.

As Bolancé and Guillén (2009) suggested, the good properties of the transformation kernel estimation of the density function is exported to the estimation of the distribution function. In this paper we explore the methodology used by Burch-Larsen et al. (2005), employing it to estimate the distribution function of the severity of operational risk of a medium- sized Savings Bank.

As Bolancé and Guillén report (2009), the estimation is done in the following way: let T (.) be a concave transformation, where y = T (x) and $Y_i=T(X_i)$, i=1…..,n are the observed transformed losses. Therefore, the kernel estimation of the transformed variable is:

$$\hat{F}_Y(y) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{y-Y_i}{h}\right) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{T(x)-T(X_i)}{h}\right) = \hat{F}_Y(T(x)) \quad (18)$$

This shows that the transformation kernel estimation of $\hat{F}_X(x)$ equals:

$$\hat{F}_x(x) = \hat{F}_{T(x)}(T(x)) \quad (19)$$

In order to obtain the transformed kernel estimation, it is necessary to determine which transformation to use, the kernel function, and to calculate the bandwidth.

In the kernel estimation of density function literature, several methods are proposed for the calculation of the bandwidth[6], however very few alternatives are analyzed in the context of the kernel function of the distribution estimation.

[6] Wand and Jones(1995).

Bolancé and Guillén (2009) propose an adaptation of the method based on the normal distribution described by Silverman (1986). This method implies the minimization of the mean integrated squared error (MISE):

$$MISE\ (\hat{F}_Y(y)) = E\left(\int \left(F_Y(y) - \hat{F}_Y(y)\right)^2 dy \right). \qquad (20)$$

The asymptotic value of MISE is known as A-MISE (Asymptotic Mean Integrated Squared Error).

By integrating the asymptotic mean square error given in expression (19) and by taking the distribution function F_Y estimate into account the A-MISE becomes:

$$n^{-1}\int F_Y(y)(1-F_Y(y))\,dy - n^{-1}h\int K(t)(1-K(t))\,dt + \frac{1}{4}k_2 h^4 \int (F''(y))^2\,dy, \qquad (21)$$

where $k_2 = \int t^2 k(t)dt \quad \int (F''(y))^2 dy = \int ((f_y'(y))^2 dy$

minimizing (20) by $\dot{h}$, overwhelming the asymptotically optimal smoothing parameter:

$$\dot{h} = \left(\frac{\int K(t)(1-K(t))\,dt}{k_2 \int (f_Y'(y))^2\,dy} \right)^{\frac{1}{3}} n^{-\frac{1}{3}} \qquad (22)$$

Silverman (1986) proposed approximating $\dot{h}$ by replacing the terms that depend on the theoretical density function with the value they would obtain if it were assumed that f is the density of a normal distribution (u, σ). By using the kernel of Epanechnikov:

$$\dot{h} = \sigma \left(900\sqrt{\pi} \big/ 35 \right)^{\frac{1}{3}} n^{-\frac{1}{3}} \qquad (23)$$

This method produces good results for almost the entire distribution, but serious problems for approximating the higher quantiles of the distribution, which in the context of measuring risk are the most important.

2.5. The double transformation kernel estimation

In this work the methodology used in Bolancé and Guillén (2009) is proposed as an alternative method. This estimation was initially applied in Bolancé, Guillén and Nielsen (2008) in the density function context and later in Bolancé and

Guillén (2009) for the estimation of a distribution function. In this article, this methodology is applied to the estimation of operational loss severity distribution. Bolancé and Guillén (2009) propose a new transformation kernel estimation, based on a double transformation, which can improve the estimation of risk measures.

The A_MISE expression given in (20) shows that to obtain the asymptotically optimal smoothing parameter it is sufficient to minimize:

$$\frac{1}{4}k_2 h^4 \int (f'(y))^2 dy - n^{-1} h \int K(t)(1-K(t))dt, \qquad (24)$$

This is minimized when the function $\int (f'(y))^2 dy$ is minimal.

The method proposed is based on transforming the variable so that its distribution is achieved by minimizing the above functional.

Terrell (1990) proves that the density of the Beta (3, 3) defined in the domain (-1, 1) minimizes $\int (f'(y))^2 dy$ among all densities with a known variance. The density functions and distribution of Beta (3, 3) are:

$$\begin{aligned} h(x) &= \frac{15}{16}(1-x^2)^2, -1 \le x \le 1, \\ H(x) &= \frac{3}{16}x^5 - \frac{5}{8}x^3 + \frac{15}{16}x + \frac{1}{2}. \end{aligned} \qquad (25)$$

Double transformation kernel estimation involves carrying out an initial processing of the data by using the distribution function of the generalized Champernowne with three parameters; Hence the transformed variable has a distribution that is located around a Uniform distribution (0,1). Subsequently, the data is transformed again by using the inverse of the Beta function (3, 3), $H^{-1}(Z_t) = Y_1, \ldots\ldots, H^{-1}(Z_n) = Y_n$.

The result of this double transformation will have a distribution close to that of Beta (3, 3). The resulting transformation kernel estimation is:

$$\hat{F}_Y(y) = \frac{1}{n}\sum_{1=1}^{n} K\left(\frac{y-Y_i}{h}\right) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{H^{-1}(T_{\hat{\alpha},\hat{M},\hat{c}}(x)) - H^{-1}(T_{\hat{\alpha},\hat{M},\hat{c}}(X_i))}{h}\right) \qquad (26)$$

The smoothing parameter h is estimated by following the methodology explained above, with the knowledge that $\int (f'(y))^2 dy = 15/7$,for Beta (3, 3), and hence by using Epanechnikov's kernel, we obtain:

$$\dot{h} = \sqrt[3]{3/n} = 3^{\frac{1}{3}} n^{-\frac{1}{3}} \quad (27)$$

3. Exploratory Analysis of the Data

The data available in this work is provided by a medium-sized Spanish Savings Bank which has compiled internal data in order to make step-by-step advances in measuring and modelling Operational Risk.

Data are daily collected by reference to the date of occurrence. The threshold for collecting the loss data in the saving bank is set at 0 Euros. Unlike other studies, where the threshold is set at 10,000 euros, 95% of the losses are lower than 542.32 Euros. Most studies apply a high threshold since information is taken from various databases. According to Carrillo (2005), the lower the threshold utilized, the more complete the information about the real distribution of the data. The data provided spans from the year 2000 to 2006. The years 2000, 2001, 2002 and 2003 contain 2, 5, 2, and 50 operational loss events, respectively, due to the bank not providing a recompilation system for operational losses in those years. However, in the year 2004 the bank provided such a system, and hence for the years 2004, 2005 and 2006 there are 6,397, 4,959 and 6,580 operational loss events, respectively.

In order to prevent a distortion of the estimation of the distribution models, data from 2000 to 2003 are disregarded. The core business of any Spanish Savings Bank is retail banking, and for this reason all the data come from this business line. Our data set is relatively small but is satisfactory enough for operational risk analysis at the level of a whole bank.

The data is adjusted according to the Consumer Price Index (CPI) to prevent distortion in the work outcome, and 2006 is taken as the baseline year.

In Table 1, all the main features on central tendency, asymmetry and tail-heaviness of the data set are given.

Table1: Descriptive statistics

Statistics	Operational risk
N	17936
Mean	254.48
Median	51.35
Standard .deviation	3602.71
Skewness coefficient	73.48
Kurtosis coefficient	6924.22

The mean is much higher than the median. This is a feature present in the skewed distributions, and is confirmed in our case by the skewness coefficients. The kurtosis coefficient also shows leptokurtic tails. The main reason for such levels of skewness and kurtosis can be found in the zero thresholds[7].

We will explore whether seasonal factors are detected in the monthly distribution of data

Table 2: Monthly amount of loss

	Year			
Month	**2004**	**2005**	**2006**	**Total**
1	66,169.1	70,337.23	38,225.93	174732.3
2	179,981.8	105,454.8	91,763.87	377200.5
3	134,809.7	249,035.4	175,849.6	559694.8
4	55,011.89	93,549.19	51,454.33	200015.4
5	168,053.5	50,433.71	106,385.2	324872.5
6	557,136.5	83,403.4	126,467.9	767007.7
7	95,412.18	558,23.84	174,745.3	325981.3
8	76,233.28	70,356.69	153,999.9	300589.8
9	80,651.07	146,472.4	190,217.6	417341
10	244,451.2	52,047.94	123,753.7	420252.9
11	84,456.23	50,670	103,134.9	238261.2
12	257,794.8	118,080.7	82,559.62	458435

[7] Higher thresholds reveal smaller kurtosis coefficients, as we have proved with the database.

Table 3: Monthly frequency of loss

	Year			
Month	**2004**	**2005**	**2006**	**Total**
1	266	428	370	1064
2	349	417	366	1132
3	538	405	490	1433
4	388	502	410	1300
5	452	418	700	1570
6	475	371	704	1550
7	439	433	655	1527
8	554	396	533	1483
9	546	412	654	1612
10	1394	416	690	2500
11	519	389	519	1427
12	477	372	489	1338

Looking at Tables 2 and 3 and Figures 1 and 2 derived from them, we note that removing the loss peak in June 2004 for the monthly amount of loss and that of October of the same year for the monthly frequency of loss, the data seems to not present seasonality.

Figure 1: Monthly amount of loss

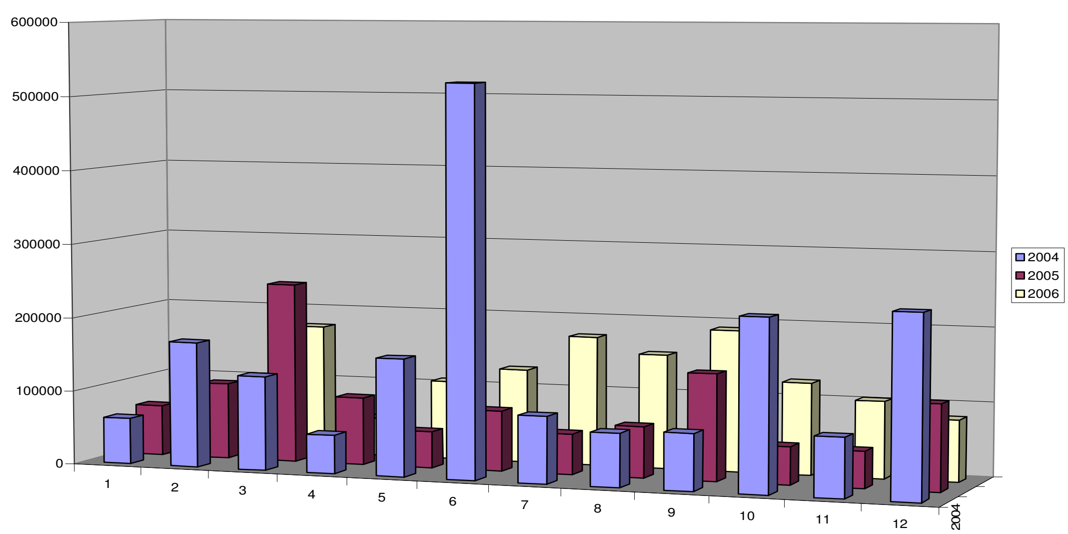


Figure 2: Monthly frequency of loss

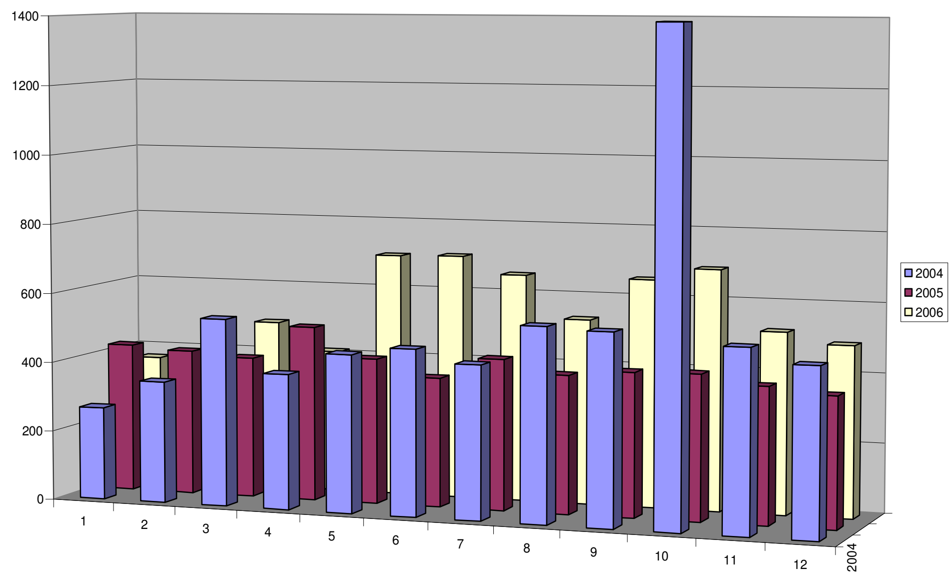
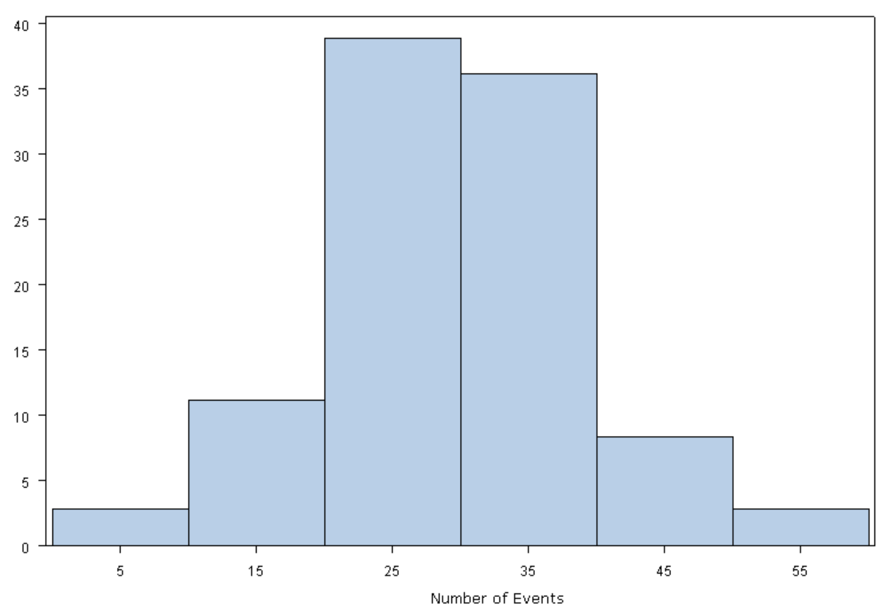


Figure 3: Daily Frequency for events above a Threshold



To test the independence of events, the graph above shows, in the abscissa, the losses that exceed a specific threshold and, in the ordinate, the number of times that these losses are exceeded. The graph approximates a binomial distribution, thereby indicating the independence of events. We chose 166 as the threshold, which is going to be used to fit the generalized Pareto distribution, and which was obtained with the method used by Reiss and Thomas (2001)[8].

Table 4: Quantiles of data

Quantile	100%	99%	95%	90%	75%	50%	25%	10%	5%	1%	0%
Estimated	352350	2474.32	540	294.3	100	50	20	10	5.15	1.15	0.03

Observing Table 4, we can notice the great difference of values between 0.99 and 1 quantiles. This figure gives an indication of the nature of the very heavy tail of these data, and this can lead to difficulty in finding the severity distribution that fits these data.

To complete the exploratory analysis of the data, we now focus our attention on the tail of the underlying distribution. Since it was anticipated in the introduction of the present work, we study how the threshold choice affects the GPD parameters estimated, especially the shape parameter, and the quantile estimation that directly affects the operational VaR.

Figure 4: Me plot and Hill plot.

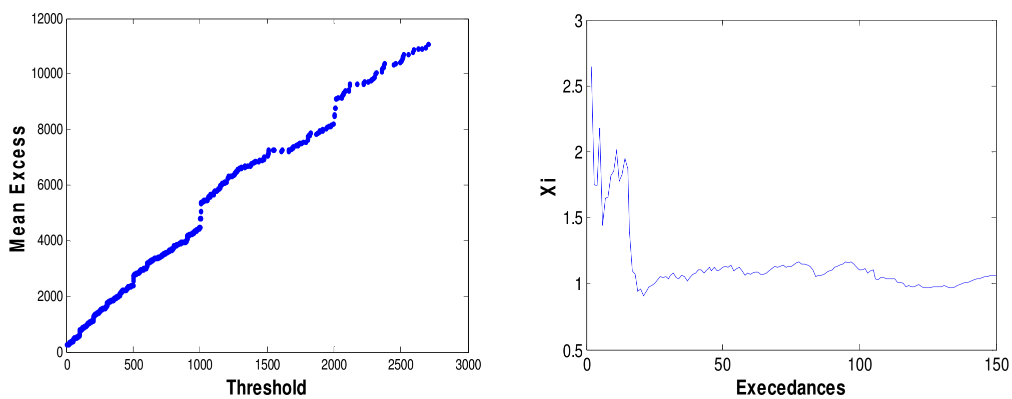


Figure 4 shows the mean excess plot (on the left) and the hill plot (on the right) respectively, obtained with the threshold chosen to fit the GPD distribution. For the abscissa of the former different thresholds are used and for the abscissa of

[8] It is an ad hoc method that the authors describe as Automatic choice of the number of extremes.

the latter different execeedances (the number of observations above the threshold).

In order to understand the problem of the choice of the threshold better, we check the effect of different thresholds on the shape parameter and on the quantile estimation.

Figure 5: Shape parameter and quantile plot.

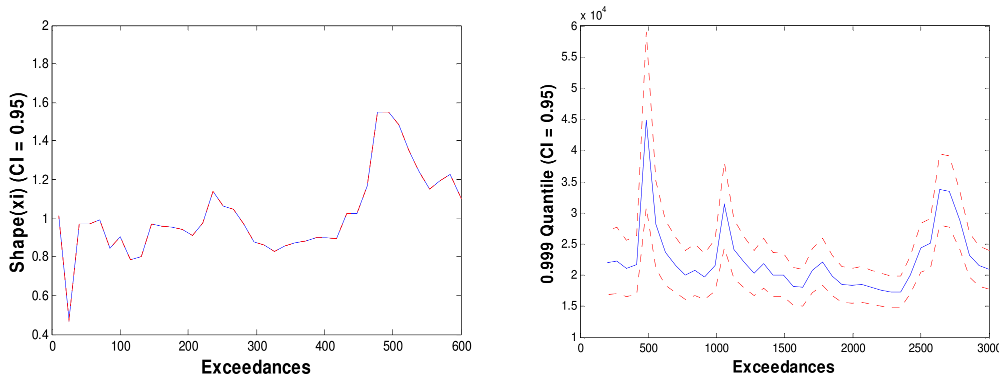


Figure 5 (on the left) plots how the estimation of the shape parameter changes depending on the number of the observations above the threshold[9], and Figure 5 (on the right) plots the estimation of the 0.999 quantile for the different number of the exceedances. Both figures highlight how the shape parameter and quantile estimations depend strongly on the threshold choice.

4. Severity Distribution

In this section we compare the distributions obtained with parametric and non-parametric estimations to approximate the severity loss distribution.

Maximum likelihood is the methodology used to estimate the parameter distributions of the parametric distributions. For the selection of threshold applied to the Generalized Pareto (GPD) the methodology used by Reiss and Thomas (2001) is employed.

[9] The number of observations above the threshold is inversely proportional to the value of the threshold.

Table 5: Parameters estimated by maximum likelihood.

Distributions	Parameters		Stand. Dev.
Weibull	α	0.59	0.36
	β	104.42	187.50
Lognormal	ɥ	3.92	1.44
	σ	1.44	1.017
GPD	ξ	0.9058	0.034
	σ	174.8585	6.16

In order to analyze the different results, we compare the results of the empirical cumulative distribution function (cdf) with each distribution resulting from the different estimation methods presented in this work, with particular attention to the tail. We divide the cdf into two parts, one part for the body that varies from 0 to 0.99 percentiles and the other part for the tail from 0.99 to 1.

To complement the visual assessment of the goodness-of-fit of different distributions, we present the result of two goodness-of-fit tests, the Cramer-Von Mises (CVM) and the Kolmogorov-Smirnov (KS).

The latter is based on the maximum vertical distance between the model proposed and the empirical distribution function. It is represented by the following formula:

$$D = \sup_{x} \left| F_n(x) - \hat{G}(x) \right| \tag{28}$$

Where D represents the distance and $\hat{G}$ the estimated cdf.

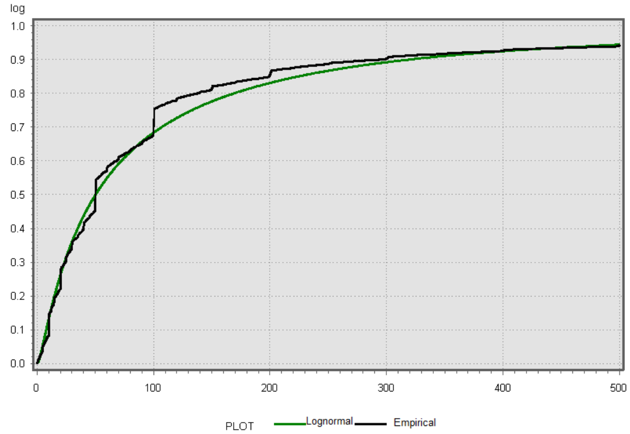
The problem with this test is that it takes into account only the maximum distance between the estimated and empirical cdf regardless of the whole setting, so we also used the Cramer-Von Mises (CVM) test.

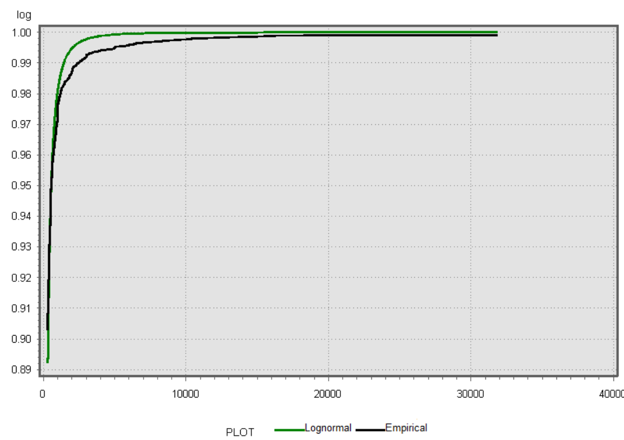
This is a measure of the square of the average distance between the data and the model being considered, with a correction in the size of the sample. The formula that represents this test is:

$$CVM = \sum (F_n(x) - \hat{G}(x))^2 + \frac{1}{12n} \tag{29}$$

Three different estimating tail quantiles for each cdf are also checked.

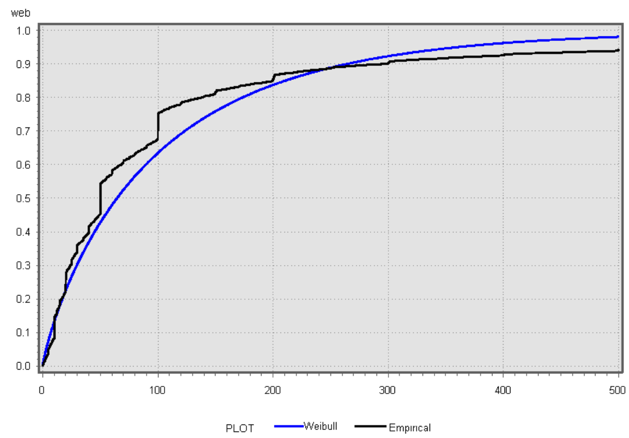
Figure 6: The cdf for the lognormal distribution compared with that of the empirical function

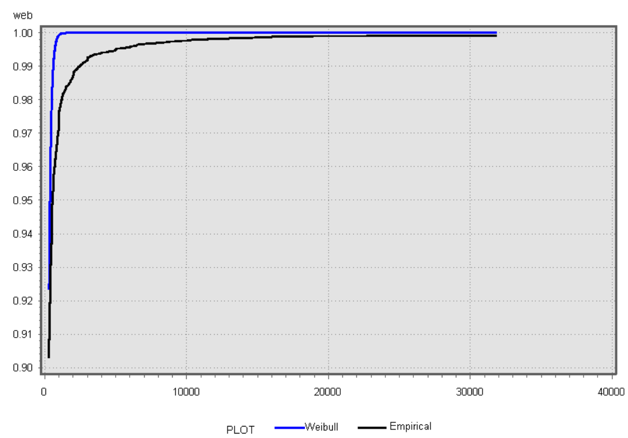




The two figures above show that the lognormal distribution fits to the body (above) of the distribution reasonably well, but does not fit the higher quantiles (below) of the distribution very well.

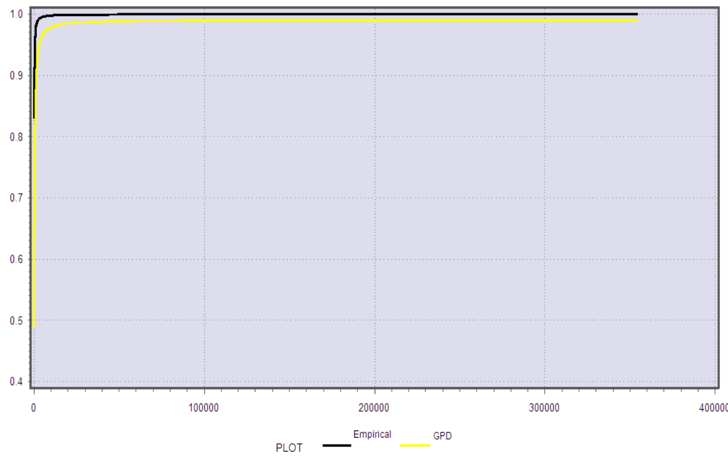
Figure 7: The cdf for the Weibull distribution compared with that of the empirical function





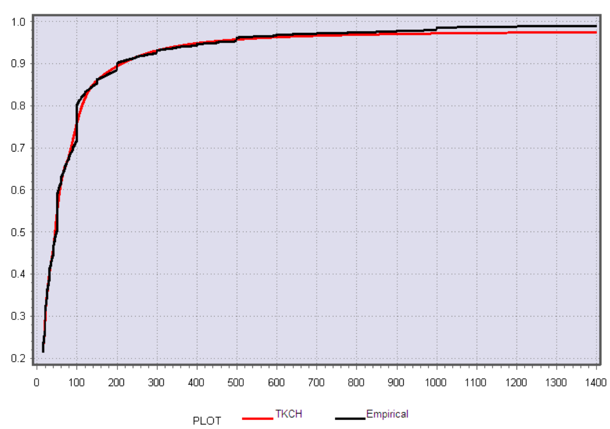
The Weibull distribution has a worse fit than the lognormal distribution in the body (above) and fits as badly as the lognormal in the tail (below).

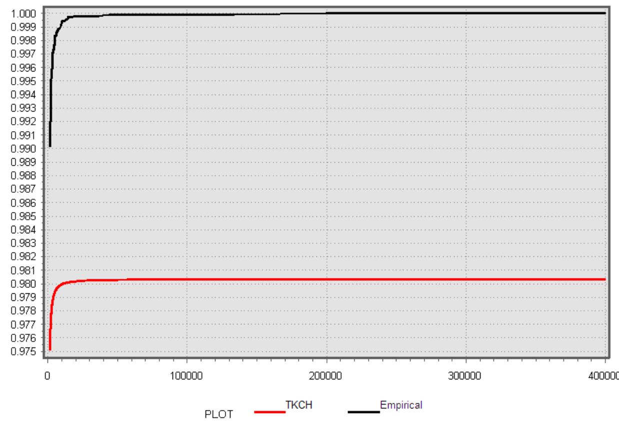
Figure 8: The cdf for Generalized Pareto Distribution compared with that of the empirical function



In the case of GPD, only the tail is considered to analyze the fit that this distribution provides. As it is possible to note from the figure, the GPD presents a better fit on the tail than the parametric distribution that does not focus the attention on the tail. The problem is the overestimation that can imply uneconomical operational VaR estimation. The GPD line plots a quite bigger amount for each quantile than the empirical line.

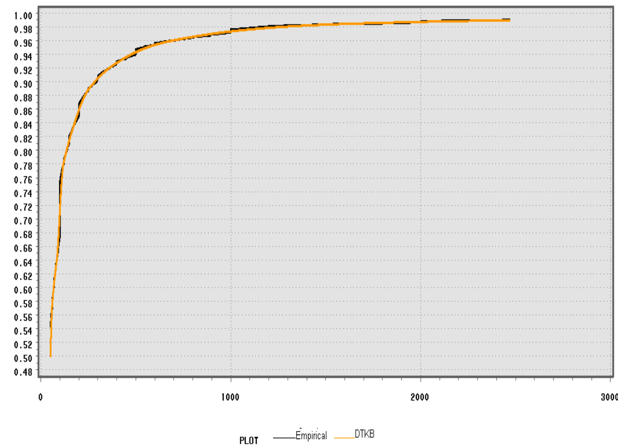
Figure 9: The cdf for the transformation kernel estimation compared with that of the empirical function

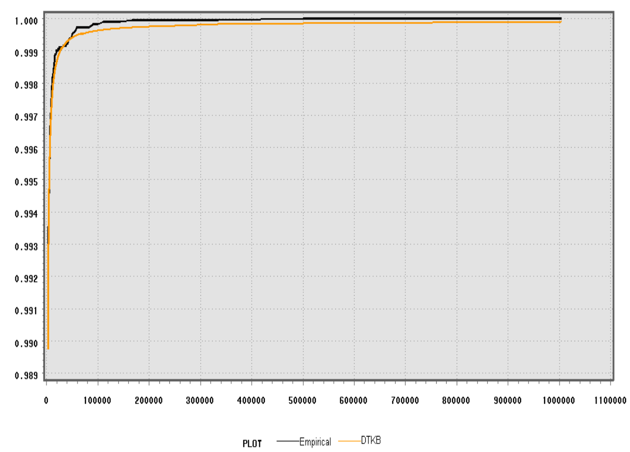




The kernel cdf estimation using the Champernowne distribution provides a good fit in the body of the distribution but has a serious problem in the tail due to the lack of information in this part of the distribution. This result is coherent with Bolancé and Guillén (2009). In the work of the authors the quantile estimation carried out by this methodology does not approximate the higher quantiles.

Figure 10: The cdf for the double transformed kernel estimation compared with that of the empirical function





As shown in Figure 10, the double transformation in the kernel estimation function significantly improves the fit of the cumulative distribution in both the body and in the tail.

Table 6: Goodness-of-fit tests and quantiles estimation

CDF	**K-S**	**CVM**	**α=0.95**	**α=0.99**	**α=0.999**
Lognormal	0.069	0.00047	533	1,427	3,718
Weibull	0.18	0.0058	529	1,098	2,246
GPD	0.036	0.00016	2895.2	12,608	10,2660
TKCH	----	-------	557	10,036	$+\infty$
DTKB	0.009	0.00007	560	2,535	27,233

GPD Generalized Pareto Distribution

TKCH Transformation Kernel Estimation with Chapernowne Distribution

DTKB Double Transformation Kernel Estimation

Table 6 shows the coefficients of the goodness-of-fit tests of the two tests mentioned above and the values produced for each cdf quantiles 0.95, 0.99, and 0.999. The lowest values of the two tests allow us to conclude that the distribution resulting from the double transformation kernel estimation provides the best fit to the data according to both goodness-of-fit tests. This good performance is coherent with the work of Bolancé and Guillén (2009).

5. Operational VaR

As a last step we analyze the performance of the non-parametric methodology, in particular the doble transformation kernel estimation, as an alternative to the approximation of the severity distribution and its subsequent use for the aggregate distribution of losses, with the aim of estimating an operational VaR. In this regard, we aggregate the severity distribution obtained with the different estimations proposed (we exclude the transformation kernel estimation because does not approximate high quantiles) with a Poisson distribution ((λ = 5978). For the calculation of VaR lognormal, weibull and DTBK we perform the following operational steps (see de Fountanelle et al, 2003):

- We simulate n years (the parameters of our work are estimated on an annual basis).
- For each year, we extract λ numbers of events from the Poisson distribution.

  For each λ events we draw an operational loss from the estimated distribution.
- Adding up the losses for each year, we obtain the aggregate distribution of losses.
- We repeat this process 100,000 times.

Finally, we sort the losses and calculate the VaR at a 95, 99 and 99.9% confidence level.

For EVT techniques, the operational VaR is estimated in the following way:

- We simulate n years (the parameters of our work are estimated on an annual basis).

- For each simulated year, we draw a frequency λ from the Poisson distribution.
- We multiply λ by the fraction of the data in the body to get λ_b and the fraction of the data in the tail to get λ_t. We draw λ b loss severities (with replacements) from the data in the body (empirical sampling) and λ_t loss severities from the estimated GPD in the tail.
- We sum all the λ_b and λ_t losses together to get the total annual loss. Finally, these steps are repeated 100,000 times.

Table 7: Results of operational VaR (0.95, 0.99, 0.999).

	Lognormal	Weibull	GPD	DTKB	Empirical
	Opvar	OpVar	OpVar	Opvar	OpVar
95^{th}	819,296,286	795,324,432	4,216,711	3,513,613	1,812,978
99^{th}	866,386,188	811,137,231	11,763,538	4,311,743	1,992,326
99.9^{th}	910,444,252	852,321,221	98,143,348	4,988,502	2,318,679

Table 7 shows the results of the aggregations of the various distribution functions. The VaR estimated with the parametric estimation tends to underestimate (lognormal, Weibull) or overestimate the real VaR. The VaR estimated with the double transformation kernel estimation seems to presents the more realistic estimation compared with the empirical estimation. These results confirm our initial hypothesis as outlined in the Introduction of this work.

Conclusions

In recent literature on operational risk, the severity loss distribution is the main topic. Numerous modelling methods have been suggested although very few work for both high-frequency small losses and low-frequency big losses. Hence, common sense suggests the estimation of a mixture of these two distributions. For small losses, distributions, such as lognormal and Weibull, are frequently used in combination with Extreme Value Theory.

Attention is then focused on an alternative one-method-fits-all approach, and we analyse the transformation kernel estimation method and the double transformation kernel estimation. The good performance of the latter in the context of operational risk severity is worthy of a special mention. In our opinion, these methodologies, especially the double transformation kernel estimation, can provide an excellent alternative for the estimation an operational risk loss severity distribution. This methodology takes into account all tail behaviour and also includes data of high-frequency small losses that form the body of the distribution.

To test the good performance of the distribution obtained from these estimation methods, we compare it with the most frequently used parametric estimation methods. The results show that in the context where the most frequently used parametric estimations are not able to fit the data or provide operational VaR economical unrealistic, the double transformation kernel estimation enables a better approximation and a more realistic operational VaR.

Bibliography

Basel Committee on Banking Supervision (2001), “On the Regulatory Treatment of Operational Risk”, Working Paper.

Bolancé, C., Guillén, M. (2009) “Estimación Núcleo Transformada para aproximar el Riesgo de Pérdida”, Investigaciones en Seguros y Gestión de Riesgos: Riesgo 2009, Cuadernos de la Fundación.

Bolancé, C., Guillén M. and Nielsen, J.P. (2008): “Transformation Kernel Estimation of Insurance Claim Cost Distributions” in Corazza, M. and Pizzi, C (Eds.) Mathematical and Statistical Methods for Actuarial Sciences and Finance, Springer, 223-231

Buch-Larsen, T., Guillén, M., Nielsen, J.P. and Bolance, C. (2005) “Kernel density estimation for heavy-tailed distributions using the Champernowne transformation”, Statistics, 39, 6, 503-518.

Carrillo Mendez, S., (2005), “Operational Risk: The Search of Reliable and Stable Economical Capital Estimation. En Fields Institute”, PRMIA Risk Management Seminar. Toronto, Canada.

Champernowne, D.G., (1952). “The Graduation of Income Distributions” *Econometrica*, 20, 4, 591-615.

Chernobai A., Rachev S., Fabozzi F., (2006), *Operational Risk: A guide to Basel 2 Capital Requirements, Models, and Analysis*, John Wiley and Son, Inc. New Jersey.

de Fontnouvelle, P., Deleus-Rueff, V., Jordan, J. and Rosengren, E., (2003), “Capital and Risk: New Evidence on Implications of Large Operational Losses”, Federal Reserve Bank of Boston, Working Paper.

Dionne, G., Dahen, H., (2007), “What about Undervaluating Operational Value at Risk in the Banking Sector?” , Working paper.

Dupuis, D.,(2004), Exceedances over High Thresholds: A Guide to Threshold Selection. Springer U.S.

Dutta, K. , Perry, J. , (2006), “An Empirical Analysis of Loss Distribution Models for Estimating Operational Risk Capital”, Federal Reserve Bank of Boston Working paper.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997), *Modeling Extremal Events for Insurance and Finance*, Springer Verlag, Berlin.

Frachot, A., O. Moudoulaud and T. Roncalli (2003). “Loss Distribution Approach in Practice,” in The Basel Handbook: A Guide for Financial Practioners, edited by Micheal Ong: Risk Books, 2004.

Gustafsson, J., Nielsen, J.P., Pritchard, P. and Roberts, D. (2006). “Quantifying Operational Risk Guided by Kernel Smoothing and Continuous credibility: A Practioner’s view”. *J*ournal of Operational Risk, 1, 1, 43-56.

Matthys G., and Beirlant J., (2003) “Estimating the Extreme Value Index and High Quantiles with Exponential Regression Models”. Statistica Sinica 13(2003), 853-880

Moscadelli, M. (2004), “The Modelling of Operational Risk: Experience with the Analysis of the Data Collected by the Basel Committee”, Technical report, Bank of Italy.

Reiss, R.D., (1981), “Nonparametric estimation of smooth distribution functions” Stand J Statist, 8, 116-119.

Reiss, R. D., and Thomas J. A. (2001). *Statistical Analysis of Extreme Values*. Birkhäuser, Basel.

Silverman, B. W., (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.

Swanepoel, J. W. H., (2005). “A new Kernel distribution function estimator based on a non-parametric transformation of the data”. Scand. J. Statist. 32,551-562.

Terrell, G.R., (1990), “The maximal smoothing principle in density estimation”, Journal of the American Statistical Association, 85, 270–277.

Wand, M.P. and Jones, M.C. (1995). Kernel Smoothing. Chapman & Hall.

Wand, P.,Marron, J. S., and Ruppert, D. (1991). Transformations in density estimation. *Journal of the Americal Statistical Association* 30, 405–417.