Using Weighted Distributions to Model Operational Risk

February 4, 2013

Pedro Corte Real
Departamento de Matemática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal (parcr@fct.unl.pt)
Lourdes B. Afonso
Centro de Matemática e Aplicações and Departamento de Matemática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal (lbafonso@fct.unl.pt)

Abstract

The quantification of operational risk has to deal with various concerns regarding data. One of the main questions is the bias in the data on the operational losses amounts recorded even if it's compiled internally. We show that it's possible, based on mild assumptions on the internal procedures put in place to manage operational losses, to estimate the parameters for the losses amounts, taking in consideration the bias that, not being considered, generates a twofold error in the estimators for the mean loss amount and the total loss amount, the former being overvalued and the last undervalued. We consider that the probability that a loss is reported and ends up recorded for analysis, increases with the size of the loss, what causes the bias in the database but, at the same time, we don't consider the existence of a threshold, above which, all losses are recorded. Hence, no loss has probability one of being recorded, in what we defend is a realist framework. We deduce the general formulae, present simulations for common theoretical distributions, estimate the impact for not considering the bias factor when estimating the value at risk and estimate the true total operational losses the bank incurred.

Key Words: Operational Risk Management, Loss Data, Bias, VaR, Applications and Case Studies.

1 INTRODUCTION

The quantification of operational risk has to deal with various concerns regarding data, much more than other types of risk which banks and insurers are obliged to manage. Several studies, at first more empirical and at present already more theoretical and mathematical supported, document several of those concerns. First of all, the lack of internal or external data on operational losses. Although this problem has, in the last years, been dealt with by researchers and practitioners, by using data collected by commercial vendors, these commercial databases still have various handicaps that, more or less, summarize the problems regarding operational loss data and, at the same time, drives the motivation to our approach to the problem of making parametric inference, using loss data statistics, in some cases aggregated data, e. g., totals or mean values.

We can summarize the main problems for operational losses data by:

a) Some of the databases reported in several papers, contain data only for big banks. Since, at least to our knowledge, there are no studies documenting that we are dealing with scale factors, meaning that, an increase in scale results in a proportional increase in operational risk, we can say that part of the industry is not represented. Another question is raised by the methods applied to compile the databases that, usually, have to depend on public disclosed losses. See, for instance de Fontnouvelle et al. (2003), where the authors compare results for two commercial databases collected this way, raising some interesting questions about the data or de Fontnouvelle et al. (2005) where some concerns about a completely different collection method and database are reported.

b) Usually these vendors can collect only data for losses that exceed some threshold, 1 million USD being common.

c) Deciding if a loss is an operational loss or not, is another problem posed to data compilers and, once decided that the loss can be classified as such, they have to define in which *business line* and *type* to classify the loss. The common classification being eight business lines, *Corporate Finance; Trading and Sales; Retail Banking; Payment and Settlement; Agency Services; Commercial Banking; Asset Management; and Retail Brokerage* and seven loss types, *Internal Fraud; External Fraud; Employment Practices and Workplace Safety; Clients, Products and Business Practices; Damage to Physical Assets; Business Disruption and System Failure; and Execution, Delivery and Process Management.*

Hence, when considering the caveats above we can say that **a)** and **b)** poses problems of bias. In the first case, we have a *structural bias* due to the large size of companies that supply the data, leaving us with a potentially biased database of institutions. For several reasons, mainly because the data is compiled from publicly available sources, only large institutions are considered, this should motivate not so large institutions to compile their own data resulting from their specific experience. In the second case, we have to deal with a confirmed biased sample of operational losses since, the vendors or the data collectors, only report data above a predefined threshold. Again, small companies will not be represented if only large losses are recorded. In the last case **c)**, we can have misclassification of operational losses, where some losses will not be reported as operational losses, or end up wrongly labeled among the line of business or loss type.

Our motivation is to propose a method to deal with the bias posed not only by the references **a)-b)** above, but also by our experience when dealing with small size insurers and banks. Our experience tells us that it is unlikely that all operational losses end up reported.

Even when the institutions have in place methods to detect and document operational losses, intending to be exhaustive and error free, not every operational loss ends us reported. There are two main reasons for that relative small losses, unless all the process is automated, will tend not to be reported. Firstly, it usually implies cumbersome work and the time used is perceived by professionals not to provide a good cost/benefit relation. Secondly, more usually than not, it implies to recognize an own or a colleagues' error. So that, we are lead to consider that there is a size bias, making more likely to report bigger losses than small losses. However, mainly due to protect the company image and reputation, even some of the largest losses can end up not being reported.

This final consideration being our leitmotif. We are lead to believe that, when dealing with loss data reported due to operational risk, we are always in the presence of a biased sample, no matter if the data comes from a commercial vendor or it is provided by internal procedures to manage operational losses. Even in the situation where there is no threshold for the losses being recorded, that is, even when the institutions try to record all operational losses, we think that the probability of a loss being reported, is still positively correlated with value of the loss, but, at the same time, not all the largest losses are reported. Meaning that, even for high thresholds, there is a chance that a loss will not be reported. The framework for this paper is that, the probability that a loss is reported and ends up recorded for analysis, increases with the size of the loss, what causes the bias in the database but, at the same time, we don't consider that a threshold exists, above which all losses are recorded and available for analysis, hence, no loss has probability one of being recorded.

We present some data, collected by a small Portuguese retail bank that, due to disclosure concerns we will not identify. For instance, in this case, the risk department estimated a probability of 1/250, for an operation to generate a operational loss and of 95%, for the loss ending up reported and documented.

2 SAMPLING FRAME AND SAMPLE

We consider that the original stochastic process we want do model is represented by the random variable *(rv)* X with a cumulative distribution function *(cdf)* $F_X(\cdot)$. In our case the *rv* will be the individual operational loss amount.

We follow the usual model and consider that this stochastic process originated a random sample of the operational losses occurred over a period (usually a year or several years), that is, $S_X = \{X_i, i = 1, ...N\}$ with the X_i independent and identically distributed (*i.i.d.*) with $F_X(\cdot)$.

Now, consider that, due to several reasons, some presented in Section 1, it is possible that not all the observations originated by X are to be registered and considered in future analysis, that is, not all the observations presented in the original sample S_X, will be available to model operational losses and for statistical inference, namely, parametric estimation. The observations available for estimation we call it (naturally) *sample* and represent by $S_Y = \{Y_j, j = 1, ...M\}$, with $M \leq N$. To the unobservable S_X, produced by the original stochastic process, we call it *sampling frame.* Here we make use of the usual denominations from sampling theory, that we will be using in our results.

Let us now suppose that, each individual loss presented in S_X has a probability, say $p_i, i = 1, ..., n$, of being recorded and, in that case, belonging to the sample S_Y, the data that is available to us to study the phenomenon.

If all the observations in S_X have the same probability of being recorded, the distribution of the Y_j would not depart from the distribution of the original stochastic process. If not, the recorded observations will not have the original distribution. In this case, the sample will have a different distribution from the sampling frame.

We suppose that the researcher of operational losses ends up with a biased sample of all the operational losses that should have been reported. The bias is originated due to the positive correlation between the loss amount and the probability of being reported.

Let us know consider that each element in the sampling frame S_X, has probability of inclusion in the sample S_Y, depending on the quality of the mechanism put in place to filter the sampling frame and on the size of the element, with largest elements having bigger probabilities. If the mechanism is perfect, all the elements in the sampling frame would be selected and end up in the sample, so that we would have no loss of information and no biased sample.

At the same time, we need a sampling scheme that takes in consideration the rarity of the largest elements, without giving probability one to all the elements above some threshold. That is, we want to put the probability of sampling the elements in S_X in perspective not only to their absolute values. For instance, if a loss of below 500.000 € is almost as common as a loss of below 1.000.000 €, we want to preserve this relative relation. On the contrary, if a loss of below 500.000 € is unlikely but below 1.000.000 € is very likely, we want to have a much higher probability to select 1.000.000 € than 500.000 €.

That is, once the sampling frame is defined, we want the sampling scheme, representing the mechanism put in place to record operational losses, to take in consideration the stochastic process origination the sampling frame and not only if a loss amount is twice another loss amount.

Let us consider that, after realization, the probability for an operational loss to be reported (or recorded, using the terminology of the probability theory) is, somehow, dependent on the quality of the mechanism put in place to record operational losses, and if the mechanism is not perfect, proportional to its likelihood.

The imperfections could arise for several reasons, for instance, due to the relative small size of some losses, that the staff don't consider worthwhile to report, due to managerial decisions, misclassification and, ultimately, because perfect control systems are difficult to implement, if at all possible.

3 WEIGHTED DISTRIBUTIONS

It's well known that, the observation of the sample, $S_Y = \{Y_1, \ldots, Y_M\} \subseteq S_X = \{X_1, \ldots, X_N\}$ will only have the same distribution as the X_i presented in the sampling frame if the sampling mechanism gives equal recording probabilities to every originally observation $X_i, i = 1, \ldots, N$.

In our model N (and of course M) is a random variable, although, depending on the sampling scheme used, the distribution of M conditional on N may be a degenerated random variable.

We propose that an approach considering a sampling scheme proportional to size and depending on the likelihood, in this case, proportional to the size of the loss, should be considered when dealing with loss data reported due to operational risk. In this framework, somehow contrary to the approach that makes the trend at the moment to deal with problems of modeling and making inference for operational risk, the Extreme Value Theory and Peeks Over Thresholds, not all the largest values have to be recorded and available to the researcher. In this case, we are not sure that all the big losses are available for study or even took part in the aggregated figures reported, e.g., total losses; mean loss, that the institution produce for accounting support.

We consider that the observations appear in the frame in a given order $\{X_1, \ldots, X_N\}$ and that the sample membership indicator, $\mathbb{I}_k$, are independent with $k = 1, \ldots, N$. The sampling scheme implies naturally that the sampling is made without replacement. The sample membership indicator are distributed relating to size according to $\mathrm{P}(\mathbb{I}_k = 1 \mid X_k) = F_X^{\xi}(x_k)$, $\xi \in [0, +\infty[$.

So, $\mathbb{I}_k \mid X_k \sim B\left(F_X^{\xi}(x)\right)$ has a Bernoulli distribution with $F_X^{\xi}(x)$ the probability of success. We can say that this is a particular case of a Poisson sampling design, with inclusion probabilities proportional-to-size, about it see, for instance, Sarndal et al. (1992).

It's possible to think of ξ as a censorship parameter (other possible analogies can be a disclosure or a quality parameter). If $\xi = 0$ (implying no censorship, total disclosure of all losses or a system so effective that all losses end up reported) we would have $\mathrm{P}(\mathbb{I}_k = 1 \mid X_k) = 1$, so that $S_Y = S_X$, and we would be in the usual situation of a random sample from $F_X(\cdot)$.

However, when $\xi > 0$, we are in the presence of some degree of censorship in our sample, making more likely that big losses are included in the sample than small losses.

The following proposition helps us in establishing the framework for our model.

Proposition 1 *Let $X_1, \ldots, X_N$ be a random sample of individual losses, with X_i independent of N a random variable with support on $\mathbb{N}$. If we consider $S_X = \{X_1, \ldots, X_N\}$ as our sampling frame (or simply frame) and apply on S_X a sampling scheme proportional-to-size with no replacement, such that, $P(\mathbb{I}_i = 1 \mid X_i = x) = F_X^{\xi}(x)$, with $i = 1, ...N$, where $F_X(\cdot)$ is the cdf of X_i and $\xi \in [0, +\infty[$ is the censorship parameter, then:*

a) Not conditional on knowledge of the frame, the inclusion variables are i.i.d. Bernoulli with $\pi = 1/(\xi+1)$ the probability of success; $B\left(1/(\xi+1)\right) = B(\pi)$, that is, $P(\mathbb{I}_i = 1) = 1/(\xi+1) = \pi$, $i = 1, ...N$.

b) Since $\sharp S_Y = \sum_X \mathbb{I}_k = \sum_{i=1}^{N} \mathbb{I}_{X_i}$, $E(\sharp S_Y \mid N) = N\pi = N/(\xi+1)$.

c) $$P(S_Y = s) = \left(\frac{1}{\xi}\right)^{\sharp s} \sum_{j \geq \sharp s} \left(\frac{\xi}{\xi+1}\right)^{j} P(N = j).$$

d) $P(X_j = x \mid \mathbb{I}_j = 1) = F_X^{\xi}(x) f_X(x)(\xi+1)$, *j=1,...N*, $\xi \in [0, +\infty[$.

Proof.

a) The independence follows from the sampling scheme. The Bernoulli distribution from

$$\begin{aligned} \mathrm{P}(\mathbb{I}_k = 1) &= \int_{\mathbb{R}} \mathrm{P}(\mathbb{I}_k = 1 \mid X = x)\mathrm{P}(X = x)dx \\ &= \int_{\mathbb{R}} F_X^{\xi}(x) f_X(x) dx = E\left(F_X^{\xi}(X)\right) \\ &= \frac{1}{\xi+1}\left[F_X^{\xi+1}(x)\right]_{-\infty}^{+\infty} = \frac{1}{\xi+1}. \end{aligned} \tag{1}$$

b) It follows directly from a).

c) Conditional on the knowledge of the frame X, we have for the probability of observing the specific samples s,

$$\mathrm{P}(Y = s \mid X) = \prod_{k\in s} \pi_k \prod_{j\in S_X - s} (1-\pi_j),$$

so that, due to the independence of the inclusion variables, we have

$$\begin{aligned} \mathrm{P}(Y = s) &= \int_{\mathbb{R}^N} \prod_{k\in s} \pi_k f_X(x_k) \prod_{j\in S_X - s} (1-\pi_j) f_X(x_j) d\prod_{i=1}^{N} x_i = \\ &= \prod_{k\in s} \int_{\mathbb{R}} \pi_k f_X(x_k) dx_k \prod_{j\in S_X - s} \int_{\mathbb{R}} (1-\pi_j) f_X(x_j) dx_j \\ &= \prod_{k\in s} \int_{\mathbb{R}} F_X^{\xi}(x_k) f_X(x_k) dx_k \prod_{j\in S_X - s} \int_{\mathbb{R}} \left(1 - F_X^{\xi}(x_j)\right) f_X(x_j) dx_j \\ &= \prod_{k\in s} \left[\frac{1}{\xi+1} F_X^{\xi+1}(x)\right]_{-\infty}^{+\infty} \prod_{j\in S_X - s} \left(1 - \left[\frac{1}{\xi+1} F_X^{\xi+1}(x)\right]_{-\infty}^{+\infty}\right) \\ &= \left(\frac{1}{\xi+1}\right)^{\sharp s} \left(1 - \frac{1}{\xi+1}\right)^{N-\sharp s} \mathbb{I}_{\{\sharp s, \sharp s+1, \ldots\}}(N) \\ &\equiv \left(\frac{1}{\xi+1}\right)^{\sharp s} \left(1 - \frac{1}{\xi+1}\right)^{N-\sharp s} \mathbb{I}_{\geq \sharp s}(N). \end{aligned}$$

Now integrating in order to N, we have:

$$\mathrm{P}(Y = s) = \sum_{j\geq \sharp s} \left(\frac{1}{1+\xi}\right)^{\sharp s} \left(1 - \frac{1}{1+\xi}\right)^{j-\sharp s} \mathrm{P}(N = j).$$

d)

$$\begin{aligned} \mathrm{P}(X_j = x \mid \mathbb{I}_j = 1) &= \mathrm{P}(\mathbb{I}_j = 1 \mid X = x)\mathrm{P}(X = x)/\mathrm{P}(\mathbb{I}_j = 1) \\ &= F_X^{\xi}(x) f_X(x)(\xi+1). \end{aligned} \tag{2}$$

From this result it follows immediately that $\mathrm{P}(X_j \leq x \mid \mathbb{I}_j = 1) = F_X^{(\xi+1)}$.

■

Proposition 2 *With the assumptions of Proposition 1, the distribution of the observations in the sample, that is, the distribution of the losses recorded, hence, the distribution of the observations available to the researcher to make inference, is a weighted distribution on $f_X(\cdot)$ with weight function $w(x) = F_X^{\xi}(x)$.*

Before we start the proof we introduce the definition of *weighted distribution*. Following Rao (1965) we have,

Definition 3 *Assume that interest is in a random variable X, with probability density function (pdf) (or probability mass function (pmf)) $f(x)$, with parameters $\theta \in \Theta$ a given parameter space. Also, assume that the values x and y are observed and recorded in the ratio of $w(x)/w(y)$, where $w(x)$ is a non-negative weight function, such that $E(w(X))$ exists. If the relative probability that x will be observed and recorded is given by $w(x) \geq 0$, then the pdf of the observed data is $f_w(x) = f(x)w(x)/\omega$, where $w(x) \geq 0$ and $\omega = \int_{\mathbb{R}} w(x) f_X(x)dx = E(w(X))$.*
The pdf $f_w(x)$ is denominated the weighted pdf corresponding to $f(x)$.

We can read the yearly work on weighted distributions in Fisher (1934). The problem of parameter estimation using non equally probable sampling scheme was first addressed by Rao (1965), Patil and Rao (1977) and Patil and Rao (1978). In these papers the authors identified various sampling situations which can be modeled using weighted distributions and calculated the Fisher information for certain exponential families, focusing primarily on $w(x) = x$, for nonnegative random variables, denominating this weighted distributions by the *size-based form* of the original distribution.

Proof. (proposition 2): By considering d) in Proposition 1 and equation (1), we've

$$\begin{aligned} \mathrm{P}(X_j = x \mid \mathbb{I}_j = 1) &= F_X^{\xi}(x) f_X(x)(\xi + 1) = \frac{F_X^{\xi}(x)}{\frac{1}{(\xi+1)}} f_X(x) \qquad (3) \\ &= \frac{F_X^{\xi}(x)}{E\left(F_X^{\xi}(X)\right)} f_X(x), \end{aligned}$$

and obviously $F_X^{\xi}(\cdot)$ is non-negative, so the conclusion follows. ■

The most common situation studied in the specialized literature deals with the size-biased weighted distribution, so that $f_w(x) = x\, f(x)/E(w(X)) = x\, f(x)/E(X)$, where X is a non-negative random variable with first order moment.

In this paper we propose that this weight function, originating the denominated *sized-biased distribution*, gives to much weight to the larger losses or, if you prefer, is to light on the smaller losses, not allowing the recording of to much of smaller losses and, at the same time, does not take in consideration the original process X for the operational losses.

The introduction of the $\xi(\geq 0)$ parameter, allows us to define in a natural way the quality of the in place mechanism to record operational losses, since we have that $E(\mathbb{I}_X = 1) = 1/(\xi + 1)$, being possible for the people involved in the process of controlling and managing the operational risk, to have a good "informed guess" for the value of ξ, for instance if $\xi = 1/2$ then $2/3$ of all the operational losses end up recorded, or even, through some specific methods to estimate the parameter ξ. For instance, by inserting in the system erroneous impacts, that should be detected and document by the control system in place, with the objective of estimating the success rate $1/(\xi+1)$. Naturally, the usual statistical inference methods can and should be applied here.

From (2), we can write $f(x) = (\xi+1)^{-1}F^{-\xi}(x)f_w(x)$, as a function of $f_w(x)$ and $F(x)$. Observing that $F_w(x) = \int_{\infty}^{x} F^{\xi}(y)(\xi+1)f(y)dy = F^{\xi+1}(x)$ we can also write $f(x)$ as a function of $f_w(x)$ and $F_w(x)$:

$$f(x) = (\xi+1)^{-1}F_w^{-\frac{\xi}{\xi+1}}f_w(x), \tag{4}$$

and write $f_w(x) = (\xi+1)F_w^{\xi/(\xi+1)}f(x)$, as a function of $f(x)$ and $F_w(x)$.

4 WEIGHTED DISTRIBUTIONS APPLIED TO MODEL OPERATIONAL RISK

In this section we will consider four distributions models for the individual operations losses amounts. The Uniform, Exponential, Pareto (type I) and the Normal model. We will deduce the impact in the parameters estimates, when using aggregate data, and not considering the bias presented in the sample, produced by a mechanism to record operational losses that is not perfect, that is by not considering a $\xi > 0$, in Proposition 1.

We will consider that the operational losses, X_i in $S_X = \{X_i, i = 1, ...N\}$, the sampling frame, have *pdf* $f(x)$ and the recorded operational losses, Y_j in $S_Y = \{Y_j, j = 1, ...M\}$, the sample available to make inference, have *pdf* $f_w(x)$. We will analyze the impact for not considering the bias presented in the sample S_Y and estimating the parameters as if the distribution is the original distribution $f(x)$. We consider both situations, when the classical theoretical model is the underling model for the weighted and non-weighted sample.

For all four distributions we will first consider that we know $f(x)$ and we want to estimate $E_w(X)$ and $V_w(X)$ and secondly we consider that we know $f_w(x)$ and we want to estimate $E(X)$ and $V(X)$. Although the Uniform model does not usually fits well to data related to losses either in banking or insurance, we present it here just to gain some insight for the ξ parameter effect and to study a distribution with limited support.

4.1 The Uniform Model

Consider two cases, the first when $X_i \sim f(x)$ is Uniform in $]a, b[$ and secondly $Y_j \sim f_w(x)$ is Uniform in $]a, b[$:

1. X_i is Uniform in $]a, b[$ so $f(x) = (b-a)^{-1}\mathbb{I}_{]a,b[}(x)$ and
   $F(x) = (x-a)/(b-a)\mathbb{I}_{]a,b[}(x) + \mathbb{I}_{[b,+\infty[}(x)$ and by (2) we have the *pdf* for Y_j:

   $$f_w(x) = \left(\frac{x-a}{b-a}\right)^{\xi}\frac{1}{(b-a)}(\xi+1)\mathbb{I}_{]a,b[}(x), \tag{5}$$

   with moments:

   $$\begin{aligned} E_w(X) &= \frac{b(\xi+1)+a}{\xi+2}, \\ V_w(X) &= b^2 - \frac{2b(b-a)}{\xi+2} + \frac{2(b-a)^2}{(\xi+2)(\xi+3)} - \left(\frac{b(\xi+1)+a}{\xi+2}\right)^2. \end{aligned}$$

2. Y_j is Uniform in $]a,b[$, so that $f_w(x)$ is Uniform in $]a,b[$ and by (4) we have the *pdf* for X_i $f(x) = (\xi+1)^{-1}\left((x-a)/(b-a)\right)^{-\xi/(\xi+1)}(b-a)^{-1}$ $\mathbb{I}_{]a,b[}(x)$ and *cdf* $F(x) = \left((x-a)/(b-a)\mathbb{I}_{]a,b[}(x) + \mathbb{I}_{[b,+\infty[}(x)\right)^{\frac{1}{\xi+1}}$ with moments:

$$\begin{aligned} E(X) &= \frac{b+(\xi+1)a}{\xi+2}, \\ V(X) &= b^2 - \frac{2b(b-a)(\xi+1)}{\xi+2} + \frac{2(b-a)^2(\xi+1)^2}{(\xi+2)(2\xi+3)} - \left(\frac{b+(\xi+1)a}{\xi+2}\right)^2. \end{aligned}$$

These results are easily obtained integrating by parts.

4.2 The Exponential Model

Using the same methodology as in Uniform distribution.

1. Consider $f(x)$ the Exponential density with parameters λ and β so that $f(x) = \beta^{-1}\exp((x-\lambda)/\beta)\mathbb{I}_{]\lambda,+\infty[}(x), \beta > 0$ and $F(x) = (1-\exp((x-\lambda)/\beta))\,\mathbb{I}_{]\lambda,+\infty[}(x)$ by (2) we have:

$$f_w(x) = (\xi+1)\left(1-\exp\left(-\frac{x-\lambda}{\beta}\right)\right)^{\xi}\frac{1}{\beta}\exp\left(-\frac{x-\lambda}{\beta}\right)\mathbb{I}_{]\lambda,+\infty[}(x), \quad (6)$$

considering $x = -(\beta\ln(y)-\lambda)$, noting that $\frac{\partial}{\partial x}B(x,y) = \int_0^1 t^{x-1}\ln(t)$ $(1-t)^{y-1}dt$, where $B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt$ is the beta function. Note that $\frac{\partial}{\partial x}B(x,y) = B(x,y)\left(\psi(x)-\psi(x+y)\right)$, being $\psi(z) = \frac{\partial}{\partial x}\ln\Gamma(x)$ the digamma function. Consider also that $\psi(n) = H_{n-1}-\gamma$ and $\psi(1) = -\gamma$ where H_n is the n-th harmonic number or in the generic form, $H_x = \int_0^1(t^x-1)/(t-1)dt$, with γ the Euler-Mascheroni constant we have:

$$\begin{aligned} E_w(X) &= \lambda + \beta H_{\xi+1}, \\ V_w(X) &= \beta^2\left(\pi^2/6 - \psi'(\xi+2)\right). \end{aligned}$$

2. Consider now that $f_w(x)$ the Exponential density with parameters λ and β by (4) we have $f(x) = (\xi+1)^{-1}\left(1-\exp\left(-(x-\lambda)/\beta\right)\right)^{-\xi/(\xi+1)}$ $\beta^{-1}\exp\left(-(x-\lambda)/\beta\right)\mathbb{I}_{]\lambda,+\infty[}(x)$. Using similar calculation we have:

$$\begin{aligned} E(X) &= \lambda + \beta H_{1/(\xi+1)}, \\ V(X) &= \beta^2\left(\pi^2/6 - \psi'\left((\xi+2)/(\xi+1)\right)\right). \end{aligned}$$

4.3 The Pareto (Type I) Model

Consider now that:

1. $f(x)$ is the Pareto density with parameters α and β so that $f(x) = \frac{\alpha}{x}\left(\frac{\beta}{x}\right)^{\alpha}\mathbb{I}_{]\beta,\infty[}(x)$ and $F(x) = \left(1-\left(\frac{\beta}{x}\right)^{\alpha}\right)\mathbb{I}_{]\beta,\infty[}(x)$ by (2) we have:

$$f_w(x) = \left(\frac{x-a}{b-a}\right)^{\xi}\frac{1}{(b-a)}(\xi+1)\mathbb{I}_{]a,b[}(x). \quad (7)$$

Considering $y = (\beta/\alpha)^\alpha$ and observing that $B(1-1/\alpha, \xi+1) = \int_0^1 y^{-1/\alpha}(1-y)^\xi dy$ the moments are:

$$\begin{aligned}
E_w(X) &= \beta(\xi+1)B(1-1/\alpha,\xi+1), \\
V_w(X) &= \beta^2(\xi+1)B(1-2/\alpha,\xi+1) - \beta^2(\xi+1)^2(B(1-1/\alpha,\xi+1))^2.
\end{aligned}$$

2. Consider now that $f_w(x)$ is Pareto with parameters α and β by (4) $f(x) = (\xi+1)^{-1}(1-(\beta/x)^\alpha)^{-\xi/(\xi+1)}\,\alpha\,x^{-1}(\beta/x)^\alpha\,\mathbb{I}_{]\beta,\infty[}(x)$
   Using similar calculation we have:

$$\begin{aligned}
E(X) &= \beta/(\xi+1)B(1-1/\alpha,1/(\xi+1)), \\
V(X) &= \beta^2/(\xi+1)B(1-2/\alpha,1/(\xi+1)) - \\
&\quad \beta^2/(\xi+1)^2(B(1-1/\alpha,1/(\xi+1)))^2.
\end{aligned}$$

4.4 The Normal Model

With the Normal model, observing that $\mathrm{erf}(z) = 2/\sqrt{\pi}\int_0^z \exp(-t^2)dt$ and $\mathrm{erfc}(z) = 1-\mathrm{erf}(z)$, let us consider the same two cases:

1. Consider $f(x)$ the Normal density with parameters μ and σ^2 using (2) we have:

$$f_w(x) = (\xi+1)\left(\frac{1}{2}\mathrm{erfc}\left(\frac{\mu-x}{\sqrt{2}\sigma}\right)\right)^\xi \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{8}$$

   Considering $-(x-\mu)/\sigma = y$ and $Z \sim N(0,1)$, observing that $\int_{-\infty}^{+\infty}(\xi+1)\left(1/2\,\mathrm{erfc}\left(-y/\sqrt{2}\right)\right)^\xi(2\pi)^{-1/2}\exp\left(-y^2/2\right) = 1$ we have:

$$\begin{aligned}
E_w(X) &= \mu + \sigma E_w(Z), \\
V_w(X) &= \sigma^2 V_w(Z).
\end{aligned}$$

2. If $f_w(x)$ follows a Normal distribution with parameters μ and σ^2, then $F_w(x) = 1/2\,\mathrm{erfc}\left((\mu-x)/(\sqrt{2}\sigma)\right)$, we have $f(x) = (\xi+1)^{-1}\left(1/2\,\mathrm{erfc}\left((\mu-x)/(\sqrt{2}\sigma)\right)\right)^{-\xi/(\xi+1)}(\sqrt{2\pi}\sigma)^{-1}$ $\exp\left(-(x-\mu)^2/(2\sigma^2)\right)$. Using the same methodology as above, with $E_{w*}(Z) = \int_{-\infty}^{+\infty} y/(\xi+1)\left(1/2\,\mathrm{erfc}\left(-y/\sqrt{2}\right)\right)^{-\xi/(\xi+1)}(\sqrt{2\pi}\sigma)^{-1}\exp\left(-y^2/2\right)dy$ we have:

$$\begin{aligned}
E(X) &= \mu + \sigma E_{w*}(Z), \\
V(X) &= \sigma^2 V_{w*}(Z).
\end{aligned}$$

5 APPLICATION

Our data was provided by a Portuguese small retail bank. We are only authorized to disclosure aggregated data and, although very inconvenient for our purposes, it is somehow a "realistic" situation because, it is very usual that we are confronted with this type of constraints when dealing with very sensitive topics.

During 2010 the bank internal reports account for a total operational loss of 4.414.000 €(rounded to the thousands). This total loss was originated by 4.700 (rounded to the hundreds) operations. So that we've a mean operational loss of 939 Euros. The risk department estimated a probability of 1/250, for an operation to generate a operational loss and of 95%, for the loss ending up reported and documented. In this case $\xi_0 = (1 - 95\%)/95\%$.

In the application we will consider only the bias effect on the classical model.

5.1 The Uniform Model

Let us consider a particular case of section 4.1 where the density of the original data being Uniform with $f(x) = 1/\theta \mathbb{I}_{]0,\theta[}(x)$, and $E(X) = \theta/2$. Using (5) in this particular case we have $f_w(x) = x^\xi/\theta^{\xi+1}(\xi+1)\mathbb{I}_{]0,\theta[}(x)$, with moments:

$$\begin{aligned} E_w(X) &= (\xi+1)/(\xi+2)\,\theta, \\ V_w(X) &= \theta^2 + 2\theta^2/(\xi+2) - 2\theta^2/\left((\xi+2)(\xi+3)\right) - \theta^2(\xi+1)^2/(\xi+2)^2. \end{aligned}$$

If we estimate the parameter θ using the method of the moments, using the sample, but with the assumption that the $Y_j \sim f(x)$ and not the weighted $Y_j \sim f_w(x)$, $j = 1, ..., M$, the true distribution, we would obtain: $\hat{\theta} = 2\overline{Y}_M$, and not $\hat{\theta}_w = (\xi+2)/(\xi+1)\overline{Y}_M$.

When we compare $E_w(X)$ with $E(X)$, we have $R = E_w(X)/E(X) = 2(\xi+1)/(\xi+2)$.

For instance, when $\xi = 1/2$, so that $\mathrm{P}(\mathbb{I}_X = 1) = 2/3$, we have that $E_w(X)/E(X) = 6/5 = 1,2$, meaning that we have an increase of 20% on the expected value of the recorded losses if 1/3 of the losses end up not being recorded according to size and to $F(x)$, that is, the expected value of a recorded loss is 20% larger than the original loss.

Considering our data, we have, $R = 1.0256$ and $\hat{\theta} = 1.878$ € versus $\hat{\theta}_w = 1.831$ €.

So, when considering a Value at Risk analysis, for a one year period with a confidence level of 1%, we would obtain for the for the 99% percentiles of the individual losses, $F^{-1}_{\hat{\theta}}(99\%) = 1.859,22$ € versus $F^{-1}_{\hat{\theta}_w}(99\%) = 1.812,69$ €.

Now, using Proposition 1 *b*), we can estimate the total loss occurred during the year. This corresponds to the total of the S_X, the sampling frame or, if preferred, the true total operational losses the bank incurred. We know that, although the mean value of individual losses is smaller when estimated using the weighted distribution, the total is bigger than the 4.414.000 € reported, due to the presence of a $\xi_0 = (1 - 95\%)/95\% (> 0)$.

Generically, we have, $E\left(\sum_{i=1}^N X_i\right) = E(N)\,E(X) = (1+\xi)E(M)\,E(X)$, so, with our data and the Uniform model, we obtain, $E\left(\sum_{i=1}^N X_i\right) = (1+\xi_0) \times 4.700 \times 1.831/2 = 4.529.440$€, estimating an increase of (115.439 €) 2.62%, in the total operational losses.

5.2 The Exponential Model

With the Exponential model, consider, without loss of generality, the particular case where $\lambda = 0$ in (6) $f_wX(x) = \beta^{-1}\exp(-x/\beta)\,\mathbb{I}_{\mathbb{R}^+}(x), \beta > 0$, with $E(X) = \beta$. Using (6) in this particular case we have $f_w(x) = (1 - \exp(-x/\beta))^\xi$
$\beta^{-1}\exp(-x/\beta)(\xi+1)\mathbb{I}_{\mathbb{R}^+}(x)$, with moments:

$$\begin{aligned} E_w(X) &= \beta H_{\xi+1}, \\ V_w(X) &= \beta^2\left(1 + \psi'(2) - \psi'(\xi+2)\right). \end{aligned}$$

When we compare $E_w(X)$ with $E(X)$, we have $R = H_{\xi+1}$.

For instance, when $\xi = 1/2$, so that $\mathrm{P}(\mathbb{I}_X = 1) = 2/3$, we have that $R = H_{1,5} = 1,28$, so that, the expected value of a recorded loss is 28% larger than the original loss.

The same reasoning used for the Uniform model, and considering our data, with $R = H_{1+\xi_0} = 1,0334$ we have, $\hat{\beta} = 939$ € versus $\hat{\beta}_w = 939/1,0334 = 908,65$ €.

So, when considering a Value at Risk analysis, for a one year period with a confidence level of 1%, we would obtain for the for the 99% percentiles of the individual losses, $F^{-1}_{\hat{\beta}}(99\%) = 4.324,25$ € versus $F^{-1}_{\hat{\beta}_w}(99\%) = 4.184,51$ €.

Estimating the true total operational losses the bank incurred, $E\left(\sum_{i=1}^{N} X_i\right) = (1+\xi_0) \times 4.700 \times 908,65 = 4.495.450$ € estimating an increase of (81.449, 3 €) 1, 85%.

5.3 The Pareto (Type I) Model

With the Pareto model, we will consider the case where $\beta = 1$ but the generalization is straightforward. $f_x(x) = \alpha/x^{\alpha+1}\mathbb{I}_{]1,+\infty[}(x), \alpha \in \mathbb{R}^+$ with $E(X) = \alpha/(\alpha-1)$, $\alpha > 1$. Using (7) in this particular case we have $f_w(x) = (1 - x^{-\alpha})^{\xi}\, \alpha x^{-(\alpha+1)}(\xi+1)\mathbb{I}_{]1,+\infty[(x)}$, with moments:

$$\begin{aligned} E_w(X) &= (1+\xi)B\left(1-1/\alpha, 1+\xi\right), \\ V_w(X) &= (\xi+1)B\left(1-2/\alpha, \xi+1\right) - (\xi+1)^2\left(B\left(1-1/\alpha, \xi+1\right)\right)^2. \end{aligned}$$

$R = (1+\xi)B\left(1-1/\alpha, 1+\xi\right)(\alpha-1)/\alpha$.

The same reasoning used for the previous models, and considering our data, we have, $R = 1,0525$, and, $\hat{\alpha} = 939/(939-1) = 1,0011$ € versus $\hat{\alpha}_w = 1,00112$ €.

So, when considering a Value at Risk analysis, for a one year period with a confidence level of 1%, we would obtain for the for the 99% percentiles of the individual losses, $F^{-1}_{\hat{\beta}}(99\%) = 99,49$ € versus $F^{-1}_{\hat{\beta}w}(99\%) = 99,49$ € .

Estimating the true total operational losses the bank incurred, we have $E\left(\sum_{i=1}^{N} X_i\right) = (1+\xi_0) \times 4.700 \times 893,86 = 4.422.240$ €, estimating an increase of (8.240, 6 €) 0, 19%.

5.4 The Normal Model

With the Normal model, we don't have any information about the standard deviation of our estimate so we will consider, without loss of generality three scenarios, the case where $\sigma^2 = \mu^2$, $\sigma^2 = (1,5\mu)^2$ and $\sigma^2 = (2\mu)^2$ in (8). $E(X) = \mu$ and $V(X) = \sigma^2$ and $f_w(x) = (\xi+1)\left(1/2\,\mathrm{erfc}\left((\mu-x)/(\sqrt{2}\sigma)\right)\right)^{\xi}$
$(\sqrt{2\pi}\sigma)^{-1}\exp\left(-(x-\mu)^2/(2\sigma^2)\right)$ with moments:

$$\begin{aligned} E_w(X) &= \mu + \sigma E_w(Z), \\ V_w(X) &= \sigma^2 V_w(Z). \end{aligned}$$

When we compare $E_w(X)$ with $E(X)$, we have $R = 1 + \sigma/\mu\, E_w(Z)$.

Using simulation we produced Table 5.4 that gives, for several values of ξ, the $E_w(Z)$ and $V_w(Z)$.

Considering our data, we have, $E_w(Z) = 0,04584$, $V_w(Z) = 0,97034$, $\hat{\mu} = 939$ and which give us the results of Table 2 for each scenario.

ξ	$E_w(Z)$	$V_w(Z)$
0	0	1
0,05	0,0436237	0,9717802
0,1	0,0845134	0,9456232
0,15	0,1230976	0,9215820
0,2	0,1595146	0,8994342
0,25	0,1940403	0,8788356
0,3	0,2267963	0,8596403
0,35	0,2579512	0,8417305
0,4	0,2876004	0,8250203
0,45	0,3159077	0,8092599
0,5	0,3429962	0,7944314
0,55	0,3689178	0,7805055
0,6	0,3937355	0,7673920
0,65	0,4175952	0,7549811
0,7	0,4405459	0,7431625
0,75	0,4626823	0,7319154
0,8	0,4840094	0,7212104
0,85	0,5045644	0,7110475
0,9	0,5244416	0,7013075
0,95	0,5436312	0,6920580
1	0,5621984	0,6831728
1,05	0,5802073	0,6746808
1,1	0,5976971	0,6665238
1,15	0,6146264	0,6587113
1,2	0,6310785	0,6511954
1,25	0,6470511	0,6439900
1,3	0,6625772	0,6370587
1,35	0,6777101	0,6303338
1,4	0,6923962	0,6238948
1,45	0,7067428	0,6176528
1,5	0,7207365	0,6115930

Table 1: $E_w(Z)$ and $V_w(Z)$ for ξ from 0 to 1,5

	$\sigma = 0,5\mu$	$\sigma = 0,75\mu$	$\sigma = \mu$
R	1,02292	1,03438	1,04584
$\hat{\mu_w}$	917,96	907,79	897,85
$\hat{\sigma^2}$	220.430	495.968	881.721
$\hat{\sigma_w^2}$	210.663	463.548	806.126
$F^{-1}_{\hat{\mu},\hat{\sigma}^2}(99\%)$	513.736	1.154.730	2.052.130
$F^{-1}_{\hat{\mu}_w,\hat{\sigma}^2_w}(99\%)$	490.994	1.079.280	1.876.230
$E\left(\sum_{i=1}^{N} X_i\right)$	4.541.490	4.491.180	4.441.970
increase of	2,9%	1,7%	0.6%

Table 2: Results for each scenario - Normal distribution

6 GRAPHICS FOR RATIOS

Figures 1 and 2 show the ratio between the expected value of a recorded loss and the original loss, for $\xi \in [0, 1]$ and for the four distributions. In Figure 2 we plot the ratio for the three coefficient of variation used in the Normal model.

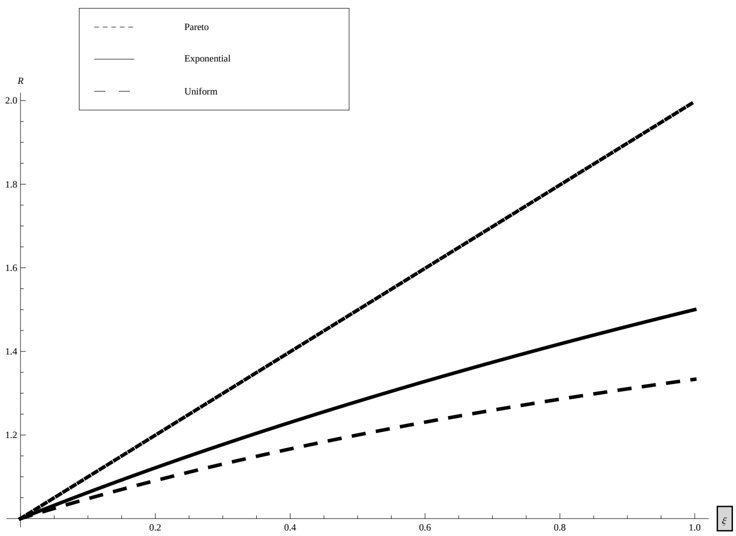

Figure 1: Ratio between the expected value of a recorded loss and the original loss for Pareto, Exponential and Uniform distributions.

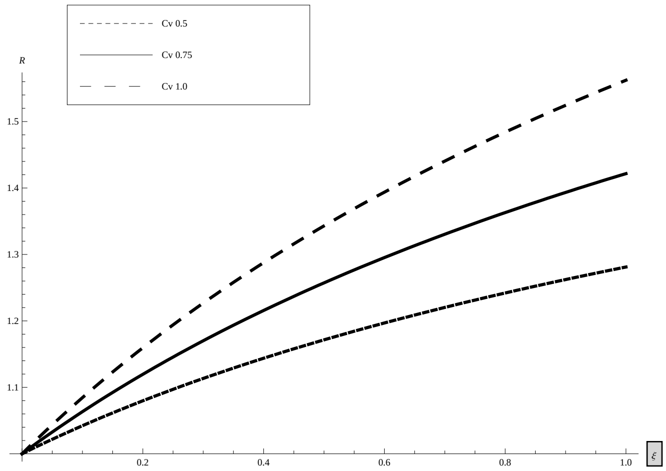

Figure 2: Ratio between the expected value of a recorded loss and the original loss, Normal Distribution for different values of σ.

7 CONCLUSIONS

Our experience tell us that, even when the institutions have in place methods to detect and document operational losses, intending to be exhaustive and error free, not every operational loss ends us reported. We are lead to believe that, when dealing with loss data reported due to operational risk, we are always in the presence of a biased sample, no matter if the data used to model the individual losses and total losses, comes from a commercial vendor or it is provided by internal procedures to manage operational losses.

Using weighted distributions we are able to consider that the probability of a loss to be reported and ends up recorded for analysis, increases with the size of the loss but, at the same time, we don't consider that a threshold exists, above which all losses are recorded and available for analysis, hence, no loss has probability one of being recorded.

Since operational risk management relies *has relied* more on qualitative approaches than on quantitative ones, more work is needed to better understand and model the exposure to operational risk. The bias presented in operational losses data, mainly due to the natural emphasis given to (public) very large losses, makes it more challenging.

Our model takes in consideration the sample bias towards the largest losses by defining a weight function functional dependent on the distribution of the original stochastic process and on the reliability recording the operational losses. In this way we can infer how the bias affects the original distribution and the estimators of the parameters.

By concentrating our attention in the expected values for the individual losses and the total losses, we learn that, for a relative high rate of success in recording operational risk losses, 95% in our example, the heavy tail distribution, that is, the Pareto's distribution, is much less affected by the bias, when estimating the parameters, than the light tail Exponential distribution.

Since our sampling scheme originates a bias towards the larger observations, if the original stochastic mechanism originating the observations has a right heavy tail distribution, the parametric estimation is less affected by the bias originated by the sampling scheme, since the observations that will not be recorded, have a bigger probability to be closer to values in the right tail. This can help to explain why the heavy tails are usually accepted as good (or nor so bad) fits to operational risk loss data.

Also, when the Normal model is considered, the increase in the variance diminishes the impact of the bias in the total operational losses estimator. The explanation being that, by increasing the volatility of the original stochastic process, we increase the probability of larger losses (relative to the mean) being recorded, hence, for an arbitrary $\xi > 0$, the soundness of the recording system, the non-recorded losses have a relative small impact in the total losses estimate.

Another important lesson to retrieve by considering the Normal model is that we should use it with extreme care. Since the support contains the negative real numbers, it would be possible for very volatile processes, to estimate a negative value for the expected value of the original stochastic process or a total mean value below the total recorded (total sample). If any of these results could be considered acceptable, this would imply that the operational losses would allow us to get some gain, for instance, by overcharging the clients.

ACKOWLEDGMENTS

Afonso's research was supported in part by CMA/FCT/UNL, under the project PEst-OE/MAT/UI0297/2011.

References

de Fontnouvelle, P., Jordan, J., and Rosengren, E. (2003). "Using Loss Data to Quantify Operational Risk" *SSRN Electronic Journal*.

de Fontnouvelle, P., Jordan, J., and Rosengren, E. (2005). "Implications of Alternative Operational Risk Modeling Techniques" Working Paper 11103, National Bureau of Economic Research.

Fisher, R. A. (1934). "The Effects of Methods of Ascertainment Upon the Estimation of Frequencies" *Annals Eugenics*, (6),13–25.

Patil, G. and Rao, C. (1977). "Weighted Distributions: A Survey of Their Application" *Applications of Statistics*, 383–405.

Patil, G. P. and Rao, C. R. (1978). "Weighted Distributions and Size-biased Sampling With Applications to Wildlife Populations and Human Families" *Biometrics*, 34(2),179–189.

Rao, C. (1965). "On Discrete Distributions Arising out of Methods of Ascertainment" *Sankhyã: The Indian Journal of Statistics, Series A*, 27(2/4),311–324.

Sarndal, C., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling.* Springer Series in Statistics.