Quantifying Operational Risk Using the Loss Distribution Approach (LDA) Model

Andrés Mora-Valencia,
School of Management,
Universidad de los Andes, Colombia.
E-mail: a.mora262@uniandes.edu.co

Wendy Zapata-Jaramillo,
Department of Finance,
EAFIT University, Colombia.
E-mail: wzapataj@eafit.edu.co

Abstract

The purpose of this article is to quantify operational risk using the Loss Distribution Approach (LDA) model. We compare two methods, one based on the Extreme Value Theory (EVT), and the second method based on an approximation named the Normal Power (NP), which consists in an Edgeworth expansion around the normal distribution. The results show that Normal Power approximation does not perform better than the Peaks Over Threshold (POT) method, based on simulation of common severities employed in operational risk. More research about expansions around distributions like gamma or lognormal is needed in order to analyze which fits better to loss distributions.

Key Words: Operational risk, Loss Distribution Approach, Normal Power, Peaks over Threshold, heavy-tailed distributions

JEL Classification: C14, C63, G2
1. Introduction

Throughout history banks and financial institutions have been affected by several operational risk events, but only during the last years, operational risk management has been subject of attention among regulators, managers, and investors following many costly and highly revealed events in the 1990s and beyond. Significant changes in the way international financial institutions estimate their regulatory capital requirements have been the focus of The Basel Committee on Banking Supervision during the last years, hence the importance of quantification and modeling this type of risk.

According to the Basel II framework, Operational Risk is defined as the risk of loss resulting from the inadequate or failed internal processes, people and systems or from external events, and every bank has to estimate its operational risk and to save the accurate amount of regulatory capital to guarantee its solvency and financial stability in the case of unexpected operational losses.

Since operational risk has been widely studied there are a variety of methods for its quantification, but there are few empirical studies trying to make operational risk measurement more standardized. In addition, it is very difficult to obtain real data in order to develop an academic research in this area. The latter makes this research a very challenging but mostly contributing topic.

Three main approaches for calculating operational risk charges are approved by the Basel II Accord, by which there was also established the minimum amount of regulatory capital. The Basic Indicator Approach (BIA); the Standardized Approach (TSA); and the Advanced Measurement Approach (AMA). One of the models most employed in the AMA is the Loss Distribution Approach (LDA).

The most sophisticated approaches are collected in the AMAs but there is a lack of a benchmark model for banks in order to measure operational risk under this technique. For this reason, banks develop their own internal models, having significant differences on their operational risk measurement methods. This research will hence seek to build a solid ground for operational risk quantification making an extensive review of the principal theory.

According to the Basel II Accord under LDA the required capital is the 99.9% Value-at-Risk (VaR) of the annual loss distribution. The 99.9% quantile can be considered a tail event, which depends on the fit of the frequency and severity distributions (fundamental components in modeling operational risk exposure) using internal and, most of times, external data. This requirement brings along with several challenges that an operational risk manager must face. In this document we shed some light on the theoretical framework that addresses this topic.

In order to address the high-quantile estimation of the loss distribution, it is common to see in practice the combination of the Loss Distribution Approach with the Extreme Value Theory (EVT). In this paper, we propose to model the heavy-tailed data of the loss distribution with the Normal Power
Approximation under the LDA framework, in order to see whether it can produce more accurate results than those obtained with the EVT.

The last proposal of Basel Committee, standardized measurement approach (SMA) has generated a vivid debate between practitioners and academics. Since SMA is not

The remainder of this paper is organized as follows. The first section approaches the literature review. The second section presents a theoretical background where an introduction to the extreme value theory (EVT) and normal power (NP) approximation are presented. The third section describes the distribution functions most commonly used in operational risk that will be used in fourth section for simulation results and application. Finally, the last section concludes.

2. Literature Review

For a thorough understanding of the operational risk framework the reader may refer to literature such as Moscadelli (2004), De Fontnouvelle et al (2004), McNeil et al (2005), Panjer (2006), and Dutta and Perry (2006).

Following the announcement of the new capital agreement, the operational risk management and its measurement models have been subject for many researches and studies made by several international scholars. Methods like Extreme Value Theory model and Bayesian networks among many others innovative techniques are commonly developed and employed to carry out the operational risk measurement.

In the last years, many practitioners have made some empirical operational risk studies that have explored one the major challenges of the operational risk quantification: The use of both light-tailed distributions like lognormal, exponential, gamma, and Weibull, along with heavy-tailed distributions such as loggamma, loglogistic, Burr, Pareto and generalized Pareto to model and quantify operational risk. Their generalized results suggest that heavy-tailed distributions fit better than light-tailed distributions for historically observed operational losses (see, for example, de Fontnouvelle et al (2004), Moscadelli (2004), and Dutta and Perry (2006)). Nonetheless, fitting heavy-tailed distributions is challenging and it also comes with several complexities.

A critical problem in need of a solution is then a high-quantile (99.9%-VaR for instance), parameter estimation for heavy-tailed distributions, and there is an extensive body of knowledge trying to serve as a statistical toolkit for tackling this problem (see, Embrechts et al (1997), Diebold et al (2001), and Degen et al (2006)).

Several books and articles of operational risk quantification review the challenges with fitting heavy-tailed distributions and some of them have been written on the application of the Extreme Value Theory (EVT) as a useful set of techniques for a quantitative risk manager as long as an optimal solution for these problems (See, Embrechts et al (1997), McNeil et al (2005)). Chavez-
Demoulin et al. (2006), review methodological techniques from the realm of EVT for modeling non-stationary operational risk data. The authors show how copula based methods can be used to engineer stress tests for dependence structures within operational risk data. The discussed techniques are advanced Peaks over Threshold (POT) modeling, the construction of dependent loss processes and the establishment of bounds for risk measures under partial information, and can be applied to other areas of quantitative risk management.

Nonetheless, some other authors highlight that the EVT should not be used blindly and suggest other methods for improving the estimation. Ergashev et al. (2013), introduce and perform a new estimation strategy for quantifying operational risk methods under the LDA framework. Their method is based on a Bayesian approach to EVT that allows incorporation of external information in order to reduce uncertainty about capital estimation when there is scarcity of internal data in order to estimate the tail behavior of the loss distribution in an accurate manner. The authors perform simulations under the Markov chain Monte Carlo approach (MCMC) and the results demonstrate that this approach clearly reduces the bias and the root mean square error of the capital estimate compared with the conventional estimation methods. With the same purpose, Nešlehová et al. (2006) highlight some issues regarding the correlation effects, the use of the EVT and the consequences of the extremely heavy-tailed data that characterize the distribution of the operational losses.

On the other hand, Makarov (2006) describes some issues concerning estimation of high quantiles and shortfalls using EVT. The author argue that EVT cannot estimate high quantiles or shortfalls precisely, since it does not yield uniform relative quantile convergence.

Chernobai and Rachev (2006) discuss the application of robust estimation methods to modeling heavy-tailed operational losses. The authors state that “Low frequency/high severity” events are an important characteristic of the operational loss data but have the same properties of outliers. The authors also explain that the application of a robust methodology to the operational loss data suggest that extreme events that account for the highest 5% of data, stands at roughly 58% to 76% of the annual aggregate expected loss and the operational risk regulatory capital charge. This paper proposes quantitative model that can provide a solution to the critics made to the Basel II framework which argue that the standards required on calculation of regulatory capital are such that this figure might even exceed the economic capital (Currie, 2005), leaving decreased availability of funds required for financial needs and investments. Although the authors highlight that the model selection is complicated by scarcity of available data along with the presence of “low-frequency/high-severity” losses, which contribute to the heaviness of the upper tail of the loss distribution.

Ergashev (2009) addresses the challenges of the heavy-tailed lognormal-gamma distribution, to estimate operational risk using the Markov chain Monte Carlo (MCMC) method and imposing prior assumptions about the unknown parameters of the model. This method is very attractive to model
operational losses, since the losses exhibit heavy tails. The author employs the quantile distance estimation method in combination with other processes. His conclusion is that prior assumptions are helpful to reduce the volatility of the estimation of the regulatory capital. Lehérisé and Renaudin (2013) employ the s-called Quantile Distance (QD) Estimation to operational risk quantification under LDA framework. A research comparing this model with the common quantification methods, which utilizes real data sets from the industry is also performed. The authors conclude that the QD method provides a useful mathematical characterization that deals with the complexities of operational risk data (heavy tail and truncation characteristics).

Miranda (2014) provides a solution for estimating the appropriate risk exposure to a particular extreme loss event over certain threshold. With this purpose, the author proposes an innovative algorithm assuming, where extreme losses follow an extreme value distribution above certain threshold. This method is based on the properties of the Generalized Pareto Distribution (GPD) and on the Hausman specification test in order to determine the threshold. The author demonstrates that the algorithm behaves very well in a simulated scenario applying the test to real risk data. The results show that despite the data restrictions inherent to the operational risk estimation process, the Hausman test is a more efficient and objective mechanism for estimating the GPD threshold than other conventional methods, and that it also might help to predict large exposures to particular relevant risks.

In the empirical operational risk studies, besides its inherent challenging heavy-tailed data, one of the major apprehension is that the uncertainty surrounding capital estimates becomes substantially larger as the loss distribution becomes heavier, which also leads to a very volatile capital estimation (See, McNeil and Saladin (1997); Mignola and Ugoccioni (2006)). Thus, other challenges come into consideration and they must be addressed.

Several numerical models, when constructing the operational loss distribution, arise trying to find an optimal solution for this problem. Jin and Ren (2010) implement recursions and Fast Fourier Transforms (FFT) for modeling the distribution function of bivariate compound random variables. The results demonstrate that the recursive method can only be applied on certain frequency distributions and it is not feasible to obtain high quantiles of the compound distributions. On the other hand, FFT method can be applied on arbitrary frequency distributions and it is very modeling-efficient.

Shevchenko (2010), also reviews some numerical models that can be used to estimate the distribution of aggregate losses. The author specifically presents and compares Monte Carlo, Panjer Recursion, and Fast Fourier Transformation (FFT) methods to estimate the aggregate loss distribution. The author concludes that the fastest method is FFT. The Panjer recursion is the easiest to implement, since it involves discretization error only. The Monte Carlo method is slower but simpler in
applications than the previous ones, and it can accurately handle multiple risks with dependence. Overall, the author states that each of the presented techniques has particular strengths and weaknesses that the investigator should be aware. The decision of the appropriate method is mainly made taking into account the specific objectives to be accomplished in the risk management area.

Other practitioners perform empirical studies with simulated data in order to tackle this issue. Horbenko et al. (2011), introduce robust procedures as “most bias-robust estimator”, “optimal mean squared error estimator” and “radius minimax estimator” and apply them on real data for the calculation of the regulatory capital employing the LDA framework. The authors model the severity of tail events by a GPD distribution, for the frequency of losses a Poisson distribution was used, and a Single-Loss Approximation for the regulatory 99.9% quantile of the total loss distribution was applied. The authors present robust diagnostic plots to evaluate the quality of their estimation and find that the robustly estimation of the Operational risk VaR is not higher than the estimate obtained by conventional methods.

Opdyke and Cavallo (2012) describe the challenges of truncation, the risks of Maximum Likelihood Estimation (MLE) and the advantages of robust statistics. Empirical simulations are used to compare MLE with two robust techniques: Optimally bias-robust estimator (OBRE) and the Cramér-von Mises estimator. The results show that OBREs present more advantages compared to MLE, when actual operational loss event data are employed. Gzyl et al. (2013), compare the performance of four numerical methods in order to determine the probability density of the total loss distribution when a model is known. One method is based on the maximum entropy principle applied to fractional moments; the second, is a probabilistic method based on moments; the third one, is the direct inversion of the Laplace transform by Fourier methods, and the last one is an approximate method based on a standard summation of densities of partial losses. Their results imply that all methods, except the second one (based on moments), obtain similar reconstructions, but the authors prefer the Maxentropic method, since “it does not need knowledge of the full Laplace transform, but only a few points of it”.

Hess (2011) evaluates the single-loss approximation, and compares it with the standard Monte Carlo simulation technique. The author shows that the VaR estimates are more precise under the single-loss model than the quantile estimates modeled by a Monte Carlo simulation. However, the author recognizes that this technique has been widely criticized by many other studies (see for instance Mignola and Ugoccioni, 2006. Carrillo et al. (2012), propose to apply the method of maximum entropy in order to reconstruct heavy-tailed distributions by splicing two different parametric densities. The authors conclude that the choice of the splicing threshold point is extremely influencing the process and that these problems need further numerical experimentation.
The Conditional Value-at-Risk (CVaR) measure has been also widely applied in the risk measurement literature, and its usefulness for modeling has been subject of research in the last decade. Pflug (2000), described its optimum characteristics, Anderson et al. (2001), firstly entered this method on the modeling of credit risk whereby the random variables where simulated by Monte Carlo Method. To date, CVaR is commonly used to model all type of risks since it captures the information on the tails in an accurate way. Feng-ge and Ping (2012) find that the CVaR measurement can provide a more accurate operational loss capital, whereby economic capital can be settled accordingly, based on the information of losses of some Chinese commercial banks.

Even though LDA is the most recommended approach, it does not capture the different cycles of the loss processes or its correlation, and it has some bias when modeling the frequency. It is proposed different solutions by using time series, e.g., an AR, ARFI or a general Gegenbauer process, where the selection criteria depend on the data set considered. Certainly, financial institutions should consider alternative approaches in the decision-making process as their measures are different to the conventional ones and they help to improve the robustness and reliability of their risks measures (Guégan and Hassani, 2013).

3. Methodology

3.1 Loss Distribution Approach (LDA)

The following definitions are based on Böcker y Klüppelberg (2005).

(1) Severity Process:

Severities \( (X_k)_{k \in \mathbb{N}} \) are positive independent and identically distributed (iid) random variables which describe the magnitude of each loss event.

(2) Frequency Process:

The amount \( N(t) \) of loss events in the time interval \([0, T]\) for \( t \geq 0 \) is random. The resulting counting process \( (N(t), t \geq 0) \), is generated by a sequence of points \( (T(n), n \geq 1) \) of non-negative random variables that satisfy:

\[
0 \leq T_1 \leq T_2 \leq \ldots \tag{1}
\]

and,

\[
N(t) = \sup\{n \geq 1 : T_n \leq t\}, \quad t \geq 0. \tag{2}
\]

(3) The severity and frequency processes are assumed to be independent

(4) Finally, the aggregate loss process:

The aggregate loss distribution \( S(t) \) at time \( t \), is given by:

\[
S(t) = \sum_{i=1}^{N(t)} X_i, \quad t \geq 0. \tag{3}
\]
The next section introduces the main concepts of extreme value theory, a technique commonly used to quantify operational risk.

### 3.2 Introduction to Extreme Value Theory (EVT)


The EVT states that the larger or smaller value of a set of values taken from the same original distribution tends to an asymptotic distribution that only depends on the tail of the original distribution. So, EVT is the study of the tails of the distributions. That is why EVT has become an important topic in the field of financial risk, and risk managers are interested in estimating probabilities of tails and quantiles of loss distributions since the financial data are heavy-tailed. EVT has been used to model catastrophic events in insurance and other financial events, such as unexpected credit losses.

The following sections are mainly based on Chapter 7 of McNeil et al. (2005), which introduces the principal parametric approaches studied in the literature about quantile estimation for heavy-tailed distributions: (i) Block Maxima method and, (ii) Peaks Over Threshold (POT) method.

1. **Block Maxima Method**

Consider an unknown and underlying distribution $F$ and a given data-sample that can be divided into $m$ blocks of size $n$. It is assumed that the extreme distribution of data is a Generalized Extreme Value (GEV) type with some unknown parameters $\xi, \mu,$ and $\sigma$. These parameters can be estimated using maximum likelihood (it is assumed that the maximum values per block are independent). But selection of block size generates a trade-off between bias and variance of the estimators. Small blocks generate bias and large blocks generate high variance in the estimation of the parameters. See example 7.12 McNeil et al. (2005) of the Block Maxima method applied to the returns the S&P500. Figure 1 graphically represents the method.

---

1 Sur la distribution limite du terme maximum d’une série aléatoire and Statistics of Extremes respectively.
ii. Peaks Over Threshold (POT) Method

The POT method is useful for large observations that exceed a fixed high threshold $u$. This method is more useful than block maxima, in practical applications, due to its more efficient use of extreme value data and from the point of view of a risk manager which is interested in the losses that exceed the threshold $u$. This approach has been studied by Smith (1989), Davison and Smith (1990) and Leadbetter (1991), who show the practical use of POT in extreme value theory (EVT) applications. Given a loss data $X_1, \ldots, X_n$ of a (unknown) distribution function $F$, a random number $N_u$ of losses exceed the threshold $u$. Then, $\tilde{X}_1, \ldots, \tilde{X}_{N_u}$ are called exceedances. For each of these exceedances the amount $Y_j = \tilde{X}_j - u$ of excess losses is calculated. The parameters are estimated using maximum likelihood estimation (assuming that the excess data are independent). The selection of the threshold generates a trade-off between bias and variance in the estimation. Very low threshold values generate bias in the estimation, while very high threshold values generate high variance in the estimation. Example 7.24 McNeil et al. (2005) applies POT to data loss of the AT&T stock. A graphical representation of the POT method is illustrated in Figure 2:

**Figure 2: Graphical Representation of the POT Method. Embrechts et al. (1997)**
Nevertheless, a disadvantage that presents this method is the selection of the threshold. As mentioned, the threshold selection leads to a tradeoff between bias and variance in the estimation of the parameters of the GPD. A very high threshold would result in high variance; while a very low threshold would result in high bias, especially when estimating $\xi$. There are several alternatives to try to solve this problem; however, there is not an optimum method to select an accurate threshold. A very common alternative is to estimate $\xi$ by employing Hill plot, which is a visual method based on the estimator with the same name. This estimator is calculated as follows:

$$
\hat{\xi}_{k,n} = \frac{1}{k} \sum_{j=1}^{k} \ln X_{n-j+1,n} - \ln X_{n-k,n}, \quad k = 1, \ldots, n-1,
$$

where $X$ represents the loss data and $n$ the total database. Then, Hill plot is:

$$
\{k, \xi_{k,n}; k = 2, \ldots, n\}.
$$

The idea is to calculate the Hill estimator for different values of $k$ and choose the estimator of $\xi$ where a flat region (plateau) is observed in the Hill plot (Reiss and Thomas 2007, p. 136).

In panel (a) and panel (b) of Figure 2.1, weekly losses of the AT&T stock are represented in percentage; while in panel (c) and panel (d), the losses of fire damage to insurance companies in Denmark. In the $x$-axis is represented by $k$ and in the $y$-axis is represented by $1/\xi$. In practice, it is suggested to choose $k$ between 10 and 50. In this case, the panel (b) and panel (d) show a flat fraction in the Hill plot.

**Figure 2.1: Hill Plot. McNeil et al. (2005, p. 228)**

However, sometimes this method can lead to erroneous estimations. In Figure 2.2, it is shown a simulation from a $g$-and-$h$ distribution with parameters $g = 0.1$ and $h = 1$ for the left side. While for
the right side \( g = 2 \) and \( h = 0.2 \). The true values of \( \xi \) would be 1 and 0.2 respectively. On the y-axis is represented \( \tilde{\xi} \).

**Figure 2.2: Hill Plot for a g-and-h Distribution. Degen et al. (2007)**

As seen on the left side, the Hill plot shows a flat region around 1; while, on the right side, the Hill plot shows a flat region around 0.7, when the real value is 0.2.

Another visual method used to estimate the quantile on EVT is the mean excess plot. However, there are problems when using this tool because some distributions (not close to the GPD) may mislead the diagnosis of the mean excess plot (Ghosh and Resnick, 2010.)

### 3.3 Normal Power Approximation

Normal Power (NP) Approximation consists of Edgeworth expansion and employs three moments; since in this article it is used the Compound Poisson to construct the loss distribution, we employ the moments of the compound distribution combined with the NP approximation.

First, we review the moments of a compound distribution:

\[
E[S] = E[N]E[X] \\
\text{var}[S] = E[N]\text{var}[X] + \text{var}[N](E[X])^2 \\
E[(S - E[S])^3] = E[N]E[(X - E[X])^3] + 3\text{var}[N]\text{var}[X]E[X] + E[(N - E[N])^3](E[X])^3 \\
\]

Where \( S \) is the aggregate loss, i.e. \( S = X_1 + \cdots + X_N \), and \( X_i \) are the iid losses independent from the frequency event \( N \). In the Simulations section, a Poisson process is assumed for the frequency part. Then, \( E[X] = \text{var}[X] = \lambda \).
NP approximation consists of Edgeworth expansion of the normal distribution, such that

\[ \Phi(s) \approx Pr \left[ \frac{S - \mu}{\sigma} \leq s + \frac{\gamma_1}{6} (s^2 - 1) + \frac{\gamma_2}{24} (s^3 - 3s) \right] \quad (7) \]

Then,

\[ \frac{S - \mu}{\sigma} = z + \frac{\gamma_1}{6} (z^2 - 1) + \frac{\gamma_2}{24} (z^3 - 3z) \quad (8) \]

where \( Z \sim N(0,1) \), \( \gamma_1 \) is the skewness and \( \gamma_2 \) is the excess kurtosis coefficient. The latter expression is useful to obtain the VaR equation for NP approximation.

The first normal power approximation (NPI) is truncated until the skewness coefficient, so

\[ \frac{S - \mu}{\sigma} = z + \frac{\gamma_1}{6} (z^2 - 1) \quad (9) \]

The second normal power approximation (NPII) is truncated until the excess kurtosis coefficient.

More details about the NP approximation can be found in Berger (1972), Seal (1977), Ramsay (1991), among others.

3.4 Value-at-Risk (VaR)

VaR is defined by

\[ \text{VaR}_\alpha = \inf \{ l \in \mathbb{R} : P(L > l) \leq (1 - \alpha) \} = \inf \{ l \in \mathbb{R} : F_l (l) \geq \alpha \} \quad (10) \]

In probabilistic terms, VaR is just a quantile of the loss distribution. VaR was developed by JPMorgan in 1994 and quickly became the standard measure of risk for regulators and risk managers. The next section briefly describes the most commonly distributions used to fit the severities in operational risk. Under POT method the VaR is quantified as:

\[ \text{VaR}_\alpha = u + \beta \left( \frac{\alpha}{N_{u}/N} \right)^{-\xi} - 1 \quad (11) \]

where \( u \) is the estimated threshold, \( N_{u}/N \) is an estimator for the survival probability, and \( \beta, \xi \) are the parameters of the GPD to be estimated. For the normal power NPI, the following relation is employed:

\[ \text{VaR}_\alpha = E[S] + \sqrt{\text{var}[S]} \left( z_\alpha + \frac{\gamma_1}{6} (z_\alpha^2 - 1) \right) \quad (12) \]

while for NPII:

\[ \text{VaR}_\alpha = E[S] + \sqrt{\text{var}[S]} \left( z_\alpha + \frac{\gamma_1}{6} (z_\alpha^2 - 1) + \frac{\gamma_2}{24} (z_\alpha^3 - 3z_\alpha) \right) \quad (13) \]

where \( Z_\alpha \) is the \( \alpha \)-quantile of a standard normal distribution, \( \gamma_1 \) is the skewness, and \( \gamma_2 \) is the excess kurtosis coefficient.

3.5 Operational Risk Distributions

This section presents some distributions commonly used in operational risk management, and be employed in the following section of simulations.
i. Pareto Distribution:
The density function of a Pareto distribution is given by

\[ f(x) = \frac{\alpha \theta^\alpha}{x^{\alpha+1}} \quad (14) \]

where \( \theta \) is the scale parameter and \( \alpha \) denotes the shape parameter. The Pareto distribution has polynomial tails with tail index equals to \( \alpha \), also called Pareto constant. In finance, the advantage of using a Pareto distribution instead of the t distribution is that the tail index in a Pareto distribution can take any positive value. While in the t distribution, the tail index only takes positive integer values. The survival function of a random variable \( X \) is given by \( P(X > x) = 1 - F(x) \), where \( F \) is the cumulative distribution function of \( X \). Thus, in finance, if \( X \) represents investment losses, \( P(X > x) \) is the probability that a loss is greater than \( x \). The survival function of Pareto is given by:

\[ 1 - F(x) = \left( \frac{\theta}{x} \right)^\alpha \quad (15) \]

ii. Loggamma Distribution:
The Loggamma distribution function is given by

\[ f(x) = \frac{\alpha^B}{\Gamma(\beta)} (\log x)^{\beta-1} x^{-\alpha-1} \quad (16) \]

Then, the survival function is

\[ 1 - F(x) \approx c x^{-\alpha} (\log x)^{\beta-1} \quad (17) \]

As noted by McNeil and Saladin (1997), if \( \beta = 1 \), a Pareto distribution is obtained. For different values of \( \beta \), it is obtained a Pareto distribution contaminated by a slow variation function given by \( (\log x)^{\beta-1} \). The larger the value of \( \beta \) the greater the contamination would be. For this reason, in the simulations, \( \beta \) values equal to 2 and 10 are employed. The Loggamma distribution function has a shape parameter \( \xi = 1/\alpha \).

iii. Lognormal Distribution
The lognormal distribution has been used to model data of severity in the insurance sector and in operational risk. This distribution has also been used as an assumption to model stocks’ prices. For example, the Black-Scholes approach used to model the price of options, suppose that the price of the underlying asset of an option follows a Lognormal distribution, which means that the logarithmic returns of this asset is normally distributed (a random variable \( X \) is lognormal distributed if its logarithm, \( Y = \log X \) is normally distributed).

iv. Weibull Distribution
In a Weibull distribution, when the tail index equals 1, this distribution is reduced to the exponential distribution, and when the tail index equals 2, it is like the Rayleigh distribution. The Weibull distribution behaves like a normal distribution when the tail index equals 3.5. It is a
stretched-exponential when the tail index is less than 1 and decays more slowly than an exponential distribution. For this reason, it is studied its shape parameter, $\tau < 1$.

v. Burr Distribution

The density function of Burr distribution is given by (Klugman et al., 2008):

$$f(x) = \frac{\alpha y \left(\frac{x}{\theta}\right)^{y-1}}{x \left[1 + \left(\frac{x}{\theta}\right)^{y}\right]^{\alpha+1}}$$  \hspace{1cm} (18)

where $\alpha$ and $\gamma$ are parameters of shape, and $\theta$ is the scale parameter. If $\theta = 1$, then the density function is:

$$f(x) = \frac{\alpha y (x)^{y-1}}{[1 + x^y]^{\alpha+1}}$$  \hspace{1cm} (19)

Thus, its cumulative distribution function is:

$$F(x) = 1 - (1 + x^y)^{-\alpha}$$  \hspace{1cm} (20)

The shape parameter $\xi = 1/\alpha \gamma$ in the Burr distribution. This distribution is widely used in the insurance field to simulate extreme losses and it has the following special cases:

- Loglogistic distribution when $\alpha = 1$
- Paralogistic distribution when $\gamma = \alpha$
- Pareto distribution when $\gamma = 1$

vi. g-and-h Distribution

The g-and-h distribution was introduced by Tukey in 1977. A random variable $X$ is g-and-h distributed if:

$$X = a + b \frac{e^{gZ} - 1}{g} e^{hZ^2/2}$$  \hspace{1cm} (21)

where $Z \sim N(0,1)$ is a standard normal random variable. The parameter $g$ controls the amount and the direction of asymmetry, while the parameter $h$ controls the amount of elongation (kurtosis). The distribution is more skewed to the right when $g$ takes high values, and positive values of $h$ produce positive elongation. The higher the value of $h$ the greater the elongation, whereas $a$ and $b$ are the parameters of location and scale respectively.

When $h = 0$, the distribution g-and-h becomes g distribution: $X = a + b \frac{e^{gZ} - 1}{g}$, corresponding to a scaled lognormal distribution. When $g = 0$, the distribution g-and-h is interpreted as $X = a + bZe^{hZ^2/2}$, referring as the h distribution. When $g = h = 0$, $X$ is normally distributed. The g-and-h presents the distribution shape parameter $\xi = h$. 
Dutta and Perry (2007) employs the g-and-h distribution to model operational risk severities; and Degen et al. (2007) find that if the losses are well modeled by a g-and-h distribution, with \( g, h > 0 \), the estimation of high quantiles using POT method usually converge very slowly and therefore its estimation by EVT would be inaccurate in particular cases.

vii. Generalized Champernowne Distribution

The density function of a generalized Champernowne distribution is given by:

\[
    f(x) = \frac{\alpha(x + c)^{\alpha-1}[(M + c)^\alpha - c^\alpha]}{[(x + c)^\alpha + (M + c)^\alpha - 2c^\alpha]^2}
\]  

(22)

This version was introduced by Buch-Larsen et al. (2005) and has been recently used in operational risk quantification (Gustafsson and Thuring, 2008, and Buch-Kromann, 2009). The reason for using this distribution is that its body is similar to the lognormal distribution and its tail to a Pareto distribution. When \( c = 0 \), it is obtained the distribution function of a Champernowne. The Champernowne distribution function presents a shape parameter \( \xi = 1/\alpha \). An advantage of this distribution is that it does not take negative values, while both the g-and-h distribution and the GPD do.

By using data from operational risk losses and the LDA method (Loss Distribution Approach), Buch-Kromann (2009) shows that the distribution g-and-h underestimates the tail of the loss, therefore, a poor estimation of the required capital can be obtained. On the other hand, the estimation based on the Generalized Champernowne distribution, overestimates the capital charge.

3.6 Data: Simulations

On the Range of practices for OpRisk AMA the BCBS paper (2009), reports the selection distribution of 42 participating banks. For frequency distribution, banks choose Poisson (93%), Negative binomial (19%) and Binomial (3%). For severity distributions, 31% of the banks choose a single severity for the whole distribution (Lognormal 33%, and Weibull 17%). 30% of the banks select one distribution for the body and another for the tail (for the body: Empirical 26%, and Lognormal 19%). For the tail: GPD 31%, and Lognormal 14%). Other distributions employed are: Gamma, g-and-h, generalized Beta, mixture of Log-normal. In the survey, banks were able to choose more than one answer per question in the distribution choice.

To compare whether the NP approximation is capable of producing more accurate results to obtain the distribution of aggregate losses than those obtained with the EVT, several simulations are performed. The objective is to analyze under which method the estimated VaR is closer to the empirical VaR. The selection of the frequency and the severity distributions to perform the simulations is based mainly on the BCBS paper mentioned above, and other papers studied in the literature review. Hence, the simulations are made with \( \lambda =100 \) for Poisson distribution, and the severity distributions (and its parameters) as shown in the tables below.
3.7 Simulation Setup:

The thresholds selections are 10% (second column) and 5% (third column) of the tail data as shown in several studies of Peaks over Threshold (POT) applications. The following algorithm indicates the steps to obtain the loss distribution.

**Step 1:** Generate \( N \sim \text{Poisson}(\lambda) \) distribution.

**Step 2:** Generate \( X_1, \ldots, X_N \) independent and identically distributed (iid) severities, according to the chosen distributions \( G \) (i.e. \( X_1, \ldots, X_N \sim \text{iid} \ G \)), and calculate

\[
S = \sum_{n=1}^{N} X_n
\]

**Step 3:** Repeat Step 1 and Step 2, \( M \) times independently to obtain \( S_i; i = 1, 2, \ldots, M \), and estimate the 99.9% VaR by the corresponding empirical quantile of the \( S_i \).

### Table 1: Simulation Results (N=1,000 Simulations)

<table>
<thead>
<tr>
<th>( \lambda = 100 )</th>
<th>POT (( u=10% ))</th>
<th>POT (( u=5% ))</th>
<th>NPI</th>
<th>NPH</th>
<th>Empirical VaR 99.9%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student’s t ( \nu = 2 (\xi = 0.5) )</td>
<td>331.00</td>
<td>345.22</td>
<td>323.24</td>
<td>876.30</td>
<td>477.55</td>
</tr>
<tr>
<td>Pareto ( \alpha = 2, \theta = 1 (\xi = 0.5) )</td>
<td>452.32</td>
<td>529.13</td>
<td>315.63</td>
<td>1,794.23</td>
<td>612.43</td>
</tr>
<tr>
<td>Pareto ( \alpha = 0.5, \theta = 1 (\xi = 2) ) million</td>
<td>3,057.50</td>
<td>1,555.48</td>
<td>2.74</td>
<td>0.35</td>
<td>1,866.22</td>
</tr>
<tr>
<td>Loggamma ( \beta = 2, \alpha = 2 (\xi = 0.5) )</td>
<td>1,126.64</td>
<td>1,180.48</td>
<td>1,105.41</td>
<td>2,564.37</td>
<td>2,637.44</td>
</tr>
<tr>
<td>Loggamma ( \beta = 10, \alpha = 2 (\xi = 0.5) ) thousand</td>
<td>1,452.83</td>
<td>1,619.54</td>
<td>169.38</td>
<td>293.06</td>
<td>2,392.70</td>
</tr>
<tr>
<td>Lognormal ( \mu = 0, \sigma = 1 (\xi = 0) )</td>
<td>290.79</td>
<td>292.12</td>
<td>263.98</td>
<td>1,405.75</td>
<td>262.48</td>
</tr>
<tr>
<td>Weibull ( \tau = 0.5 (\xi = 0) )</td>
<td>381.79</td>
<td>380.78</td>
<td>323.22</td>
<td>922.08</td>
<td>404.03</td>
</tr>
<tr>
<td>Burr ( \alpha = 2, \gamma = 2, \theta = 1 (\xi = 0.25) )</td>
<td>114.46</td>
<td>115.02</td>
<td>115.43</td>
<td>1,142.17</td>
<td>116.08</td>
</tr>
<tr>
<td>Burr ( \alpha = 0.5, \gamma = 1, \theta = 1 (\xi = 2) ) million</td>
<td>6,341.19</td>
<td>4,196.58</td>
<td>1.12</td>
<td>1,853.80</td>
<td>1,534.23</td>
</tr>
<tr>
<td>g-h ( g=1, h=2 (\xi = 2) ) million</td>
<td>2,758.37</td>
<td>2,970.77</td>
<td>590.03</td>
<td>0.24</td>
<td>1,325.65</td>
</tr>
<tr>
<td>g-h ( g=0.1, h=2 (\xi = 2) ) million</td>
<td>445.18</td>
<td>260.84</td>
<td>34.69</td>
<td>2,418.37</td>
<td>182.26</td>
</tr>
<tr>
<td>g-h ( g=2, h=0.2 (\xi = 0.2) )</td>
<td>30,344.22</td>
<td>29,307.02</td>
<td>1,392.09</td>
<td>44,201.94</td>
<td>16,995.98</td>
</tr>
<tr>
<td>Champernowne ( m=1, c=1/4, \alpha = 5 (\xi = 0.2) )</td>
<td>146.29</td>
<td>146.47</td>
<td>146.39</td>
<td>2,311.86</td>
<td>148.37</td>
</tr>
<tr>
<td>Champernowne ( m=1, c=1/4, \alpha = 0.5 (\xi = 2) ) million</td>
<td>2,847.95</td>
<td>2,185.56</td>
<td>0.11</td>
<td>6.81</td>
<td>368.24</td>
</tr>
</tbody>
</table>

The Table 1 shows that for a small number of simulations (N=1,000), the calculated value under each of the chosen methodologies are not similar to the required 99.9% empirical VaR. Nevertheless, POT methodology seems to be closer in most of the distributions, especially POT calculated using a
5%-threshold of tail data. NPII seems to overestimate the risk in most cases, especially in the g-and-h distribution.

Table 2: Simulation Results (N=10,000 Simulations)

<table>
<thead>
<tr>
<th>Distribution</th>
<th>POT (u=10%)</th>
<th>POT (u=5%)</th>
<th>NPI</th>
<th>NPII</th>
<th>Empirical VaR 99.9%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student’s t ν=2 (ξ = 0.5)</td>
<td>435.44</td>
<td>450.43</td>
<td>447.86</td>
<td>331.32</td>
<td>431.28</td>
</tr>
<tr>
<td>Pareto α = 2, θ=1 (ξ = 0.5)</td>
<td>458.38</td>
<td>478.23</td>
<td>361.90</td>
<td>382.79</td>
<td>498.13</td>
</tr>
<tr>
<td>Pareto α = 0.5, θ=1 (ξ = 2) million</td>
<td>12,519.62</td>
<td>12,559.54</td>
<td>618.04</td>
<td>4,063.02</td>
<td>7,579.50</td>
</tr>
<tr>
<td>Loggamma β=2, α=2 (ξ = 0.5)</td>
<td>1,403.66</td>
<td>1,478.90</td>
<td>814.19</td>
<td>3,229.53</td>
<td>1,405.72</td>
</tr>
<tr>
<td>Lognormal μ = 0, σ = 1 (ξ = 0)</td>
<td>271.59</td>
<td>272.51</td>
<td>267.06</td>
<td>273.17</td>
<td>267.98</td>
</tr>
<tr>
<td>Weibull λ = 0.5 (ξ = 0)</td>
<td>1,417.98</td>
<td>1,457.41</td>
<td>381.43</td>
<td>426.97</td>
<td>1,324.44</td>
</tr>
<tr>
<td>Burr α=2, γ=2, θ=1 (ξ = 0.25)</td>
<td>113.65</td>
<td>113.74</td>
<td>111.28</td>
<td>114.29</td>
<td>113.62</td>
</tr>
<tr>
<td>Burr α=0.5, γ=1, θ=1 (ξ = 2) million</td>
<td>9,666.24</td>
<td>9,934.91</td>
<td>104.46</td>
<td>107.61</td>
<td>2,602.04</td>
</tr>
<tr>
<td>g-h g=1, h=2 (ξ = 2) million</td>
<td>8,150.31</td>
<td>5,910.85</td>
<td>19,147.56</td>
<td>5,576.38</td>
<td>11,563.37</td>
</tr>
<tr>
<td>g-h g=0.1, h=2 (ξ = 2) million</td>
<td>1,916.47</td>
<td>2,332.40</td>
<td>15,118.57</td>
<td>100.31</td>
<td>1,490.56</td>
</tr>
<tr>
<td>g-h g=2, h=0.2 (ξ = 0.2)</td>
<td>12,412.93</td>
<td>11,020.18</td>
<td>29,544.47</td>
<td>21,432.26</td>
<td>11,499.61</td>
</tr>
<tr>
<td>Champernowne m=1, c=1/4, α =5 (ξ = 0.2)</td>
<td>149.88</td>
<td>149.89</td>
<td>148.50</td>
<td>151.27</td>
<td>149.57</td>
</tr>
<tr>
<td>Champernowne m=1, c=1/4, α =0.5 (ξ =2) million</td>
<td>4,496.28</td>
<td>3,737.57</td>
<td>502.74</td>
<td>127.75</td>
<td>3,206.83</td>
</tr>
</tbody>
</table>

Table 2 exhibits mixed results, POT methodology is closer to the 99.9% empirical VaR with most of the distributions than the other methods, but in this case the POT calculated using a 10%-threshold of the tail data. This is more consistent with the literature because the 10%-threshold includes more data to calculate de VaR. In this scenario, NPI overestimates the risk outstandingly by employing the g-and-h distribution compared to the empirical VaR.

Table 3: Simulation Results (N=100,000 Simulations)

<table>
<thead>
<tr>
<th>Distribution</th>
<th>POT (u=10%)</th>
<th>POT (u=5%)</th>
<th>NPI</th>
<th>NPII</th>
<th>Empirical VaR 99.9%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student’s t ν=2 (ξ = 0.5)</td>
<td>434.11</td>
<td>452.27</td>
<td>351.71</td>
<td>572.16</td>
<td>442.57</td>
</tr>
<tr>
<td>Pareto α = 2, θ=1 (ξ = 0.5)</td>
<td>488.00</td>
<td>509.74</td>
<td>400.72</td>
<td>3,848.90</td>
<td>512.6</td>
</tr>
<tr>
<td>Pareto α = 0.5, θ=1 (ξ = 2) million</td>
<td>11,765.97</td>
<td>11,005.97</td>
<td>39,366.11</td>
<td>55,136.68</td>
<td>10,176.69</td>
</tr>
<tr>
<td>Loggamma β=2, α=2 (ξ = 0.5)</td>
<td>1,546.97</td>
<td>1,609.39</td>
<td>1,805.85</td>
<td>3,011.02</td>
<td>1,623.625</td>
</tr>
<tr>
<td>Loggamma β=10, α=2 (ξ=0.5) thousand</td>
<td>2,299.59</td>
<td>2,393.42</td>
<td>1,967.74</td>
<td>8,020.68</td>
<td>2,435.66</td>
</tr>
</tbody>
</table>
Table 3 shows that with a greater number of simulations (N=100,000; see McNeil et al., 2005), POT methodology is closer to the 99.9% empirical VaR with most of the distributions, than the other methods. In this scenario, NPII overestimates the risk in most cases compared to the empirical VaR.

4. Results and Discussion

Since it is not easy to obtain actual data of operational risk losses, based on Warnung and Temnov (2008) results, a comparison between the VaR estimations using EVT and Normal Power Approximation are performed.

Temnov and Warnung (2008) compare three methods to obtain the distribution of aggregate losses. Database comes from a financial institution with external data from a group of banks. The authors estimate 99.9%-VaR employing three methods: Monte Carlo simulation (MC), Fast Fourier Transforms (FFT) and Credit Risk + methodology (CRP). For each business line, a simulation with 5,000 replicates data on the severity and frequency of Temnov and Warnung (2008, Table 1) is performed. After obtaining the loss distribution, VaR is calculated at 99.9% through POT and normal power methods: NPI and NPII, compared with the average of the three results from Temnov and Warnung.

Table 4: Comparison of 99.9% VaR for Different Methods

<table>
<thead>
<tr>
<th>Line 2</th>
<th>M-C</th>
<th>FFT</th>
<th>CRP</th>
<th>Average</th>
<th>POT</th>
<th>NPI</th>
<th>NPII</th>
</tr>
</thead>
<tbody>
<tr>
<td>67.4</td>
<td>68.34</td>
<td>68.29</td>
<td>68.01</td>
<td>47.17</td>
<td>47.78</td>
<td>54.28</td>
<td></td>
</tr>
<tr>
<td>33.5</td>
<td>32.33</td>
<td>32.60</td>
<td>32.81</td>
<td>33.51</td>
<td>22.68</td>
<td>14.70</td>
<td></td>
</tr>
<tr>
<td>26.0</td>
<td>27.30</td>
<td>27.26</td>
<td>26.85</td>
<td>28.49</td>
<td>29.44</td>
<td>32.31</td>
<td></td>
</tr>
<tr>
<td>108.5</td>
<td>110.17</td>
<td>110.31</td>
<td>109.66</td>
<td>138.97</td>
<td>193.416</td>
<td>246.57</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>235.4</td>
<td>238.14</td>
<td>238.46</td>
<td>237.33</td>
<td>248.14</td>
<td>293.31</td>
<td>347.86</td>
</tr>
</tbody>
</table>

POT methodology has been calculated using a 10%-threshold of tail data. NP approximation is calculated based on equations (12) and (13). The results of the second to the third column of the Table
4 are obtained from Temnov and Warnung (2008). Assuming comonotonicity (i.e. perfect dependence among business lines), POT methodology is close to the average VaR of the methods, while NPI and NPII overestimate risk compared to the average.

5. Conclusions and Recommendations

Operational risk has been widely studied and there are a variety of methods for its quantification, but there are few empirical studies trying to make operational risk measurement more standardized, except for Peters et al. (2016), who propose a way to standardize AMA, which mainly consists on a hybrid LDA model with factor regression components. The most employed technique in financial industry (especially, big banks) is the loss distribution approach (LDA). However, different VaR estimations are obtained depending on the approximation subject of analysis.

POT method has good results not only in market and credit risk, but also in operational risk as shown in this article. It is also noted that Normal Power approximations do not perform better than POT method and that other approximations with other expansions need to be analyzed.

Assuming perfect dependence among business lines, POT methodology is close to the empirical VaR of the methods, while NPI and NPII overestimate risk compared to the empirical VaR. Another approximation proposed for a future research is a Laguerre expansion around the gamma distribution.

References


