Anales del Instituto de Actuarios Españoles, 4ª época, 26, 2020 / 181-224

ADVANCED MODEL OF CALCULATION OF CAPITAL FOR OPERATIONAL RISK: TOP-DOWN APPROACH [1]

MODELO AVANZADO DE CÁLCULO DE CAPITAL POR RIESGO OPERACIONAL: ENFOQUE "TOP-DOWN"

Estefanía González Carbonell *

Universidad Carlos III de Madrid. Master in Actuarial and Financial Sciences

Reception date: September 3rd 2020
Acceptance date: November 24th, 2020

Abstract

The Entities are exposed to disruptive events derived from operational risk. To provide coverage against this risk, an advanced model has been developed to quantify the Economic Capital required to cover losses due to the said risk. The model is based on the *Loss Distribution Approach (LDA)*, the *Monte Carlo simulation* and the *Value at Risk* measure, using only historical internal loss data. The model uses a "*top-down*" approach which consists in the calculation of the total Capital in the first instance and later the disaggregation of it among the Basel Cells. The model proves to be robust by capturing the Entity's true risk profile, solving the problem of data scarcity and achieving the maximum level of ex post granularity.

Keywords: Operational Risk, Loss Distribution Approach, Capital at Risk, Risk Appetite, Value at Risk

---

[1] This Paper is based on the Master's Thesis presented by the author and whose directors were the professors José Miguel Rodríguez-Pardo del Castillo and Jesús Ramón Simón del Potro at Universidad Carlos III de Madrid.

* Author for correspondence: estefania.gc@hotmail.com

DOI: 10.26360/2020_8

Resumen

Las Entidades están expuestas a eventos disruptivos derivados del riesgo operacional. Para proporcionar una cobertura frente a este riesgo, se ha desarrollado un modelo avanzado que permite cuantificar el Capital Económico necesario para cubrir las pérdidas por dicho riesgo. El modelo se basa en el enfoque de Distribución de Pérdidas Agregadas (LDA), en la simulación de Monte Carlo y en la medida Valor en Riesgo, utilizando únicamente datos de pérdida interna histórica. El modelo utiliza un enfoque "top-down" que consiste en calcular el Capital total en primera instancia y posteriormente desagregarlo entre las Celdas de Basilea. El modelo demuestra ser robusto capturando el verdadero perfil de riesgo de la Entidad, resolviendo el problema de escasez de datos y logrando el máximo nivel de granularidad a posteriori.

Palabras clave: Riesgo Operacional, Enfoque de Distribución de Pérdidas Agregadas, Capital en Riesgo, Apetito de Riesgo, Valor en Riesgo

1. Introduction

Banks are exposed to disruptive events such as failures in processes, personnel, incidents that damage or make the bank's facilities, telecommunications or information technology infrastructures inaccessible, or external events such as terrorist attacks and natural disasters that affect the human resources. All of these events are attributable to operational risk and can cause significant financial losses for the bank, as well as broader disruptions to the financial system (Basel Committee on Banking Supervision, 2011).

To provide coverage against this operational risk, an advanced model has been developed to quantify the Economic Capital required to cover losses due to said risk and the Risk Appetite for the coming year using a "*top-down*" approach. The model is based on the Loss Distribution Approach (LDA), the Monte Carlo simulation and the Value at Risk measure.

This Paper provides a summary of the Master's Thesis of González (2020). In the last, the Basel II framework has been addressed and the advanced model is described in depth.

This Paper is structured in seven sections. An introduction is made in this section, providing the scientific contribution of the model, mentioning the approach used and indicating the main underlying assumptions. Section *2* indicates the procedure to obtain the Capital at Risk and the Risk Appetite due to operational risk for next year. Section *3* presents the methodology to model Severity and Frequency, estimate the Aggregate Loss Distribution and obtain the mentioned risk measures. Section *4* describes the data sample used and a modelling study is carried out. In section *5*, the implementation of the model is put into practice. In section *6,* an evaluation of the results is carried out. Finally, section *7* contains the most relevant conclusions and the implications of the model.

The scientific contribution of this Paper resides in the fact that the proposed model provides solutions to the problems that present the current methodology used by prestigious large banks which is based on a bottom-up approach. This is detailed in *Table 1.*

Table 1. Solutions of the present model, which uses an advanced top-down approach, to the advanced bottom-up approach currently used by prestigious large banks.

Problems of the advanced bottom-up approach	Solutions of the proposed model
The external loss data is of a totally different nature since they belong to other financial institutions, so it does not reflect the risk profile of the Entity in question and the scenarios are invented situations, which implies that both the Severity and Frequency data are unreliable, making it unclear to what extent they feed or contaminate the sample.	It will be based solely on internal loss data of the Entity, relegating the use of external data and scenarios, in order to better capture its risk profile. Internal loss data is the most relevant entry, as it is clearly linked to the Entity's current business activities, technological processes and risk management procedures.
The ex-ante segmentation by Risk Unit is not consistent because it makes the size of some data samples insufficient for a distribution to be fitted without falling into overfitting problems or the like, giving rise to unreliable estimates.	Segmentation will not be carried out ex-ante. On the contrary, the internal loss data will be modelled jointly to obtain the Capital at Risk and Risk Appetite in the first instance and afterwards these measures will be disaggregated among the Cells of Basel.

Source: Own elaboration

Therefore, as seen in *Table 1*, the model captures the true risk profile of the entity by relegating the use of external data and scenarios, solving the problem of data scarcity by modelling jointly all data and achieving the maximum level of ex post granularity by making a segmentation ex-post.

Being the investigation problem diagnosed, the approach used for the development of the present model is a top-down approach, which consists of the modelling of all the historical internal loss data jointly in order to obtain the total Capital at Risk and total Risk Appetite at the entity level in the first instance, and afterwards the disaggregation of these measures among the Basel Cells presented by the Entity through an empirical approach, being each Cell a combination of a Business Line and a Risk Type. This categorization forms an array with 56 data Cells, showed in *Figure 1*, and the eight Business Lines (BL) and seven Risk Types (RT) are indicated in *Table 2*.

BL1_RT1	BL1_RT2	BL1_RT3	BL1_RT4	BL1_RT5	BL1_RT6	BL1_RT7
BL2_RT1	BL2_RT2	BL2_RT3	BL2_RT4	BL2_RT5	BL2_RT6	BL2_RT7
BL3_RT1	BL3_RT2	BL3_RT3	BL3_RT4	BL3_RT5	BL3_RT6	BL3_RT7
BL4_RT1	BL4_RT2	BL4_RT3	BL4_RT4	BL4_RT5	BL4_RT6	BL4_RT7
BL5_RT1	BL5_RT2	BL5_RT3	BL5_RT4	BL5_RT5	BL5_RT6	BL5_RT7
BL6_RT1	BL6_RT2	BL6_RT3	BL6_RT4	BL6_RT5	BL6_RT6	BL6_RT7
BL7_RT1	BL7_RT2	BL7_RT3	BL7_RT4	BL7_RT5	BL7_RT6	BL7_RT7
BL8_RT1	BL8_RT2	BL8_RT3	BL8_RT4	BL8_RT5	BL8_RT6	BL8_RT7

Figure 1. Basel Cells of the entity. Source: Own elaboration

Table 2. Business Lines and Risk Types of Basel II

Business Line (BL)	Risk Type (RT)
BL1. Corporate Finance	RT1. Internal Fraud
BL2. Trading and Sales	RT2. External Fraud
BL3. Retail Banking	RT3. Employment Practices and Workplace Safety
BL4. Commercial Banking	RT4. Clients, Products and Business Practice
BL5. Payment and Settlement	RT5. Damage to Physical Assets
BL6. Agency Services	RT6. Business Disruption and System Failure
BL7. Asset Management	RT7. Execution, Delivery and Process
BL8. Retail Brokerage	Management

Source: Own elaboration

The main underlying assumptions of the model are two: i) Frequency and Severity are two independent sources of randomness, and ii) the operational losses are considered independent and identically distributed (Frachot, Georges & Roncalli, 2001).

The model is based on the equation *(1.1)* that relates the total loss, the Severity and the Frequency, which defines the total loss as a random sum of the losses:

$$S = \boldsymbol{X}_1 + \boldsymbol{X}_2 + \cdots + \boldsymbol{X}_N = \sum_{n=0}^{\boldsymbol{N}} \boldsymbol{X}_n \qquad (1.1)$$

where

- $\boldsymbol{N}$*:* random variable representing the number of risk events in the time interval [0, t] (Frequency of events);
- $\boldsymbol{X}_n$*:* random variable that expresses the amount of loss for a certain event (Severity of the loss).

The model proves to be robust by capturing the Entity's true risk profile, solving the problem of data scarcity, achieving the maximum level of ex post granularity and providing a Capital figure with a notable degree of conservatism while improving the standard formula.

2. Model Design

This section details the procedure to be followed to obtain the Capital at Risk, Risk Appetite, Expected Loss and Unexpected Loss due to operational risk for next year. A flow chart is provided in *Figure 2.*

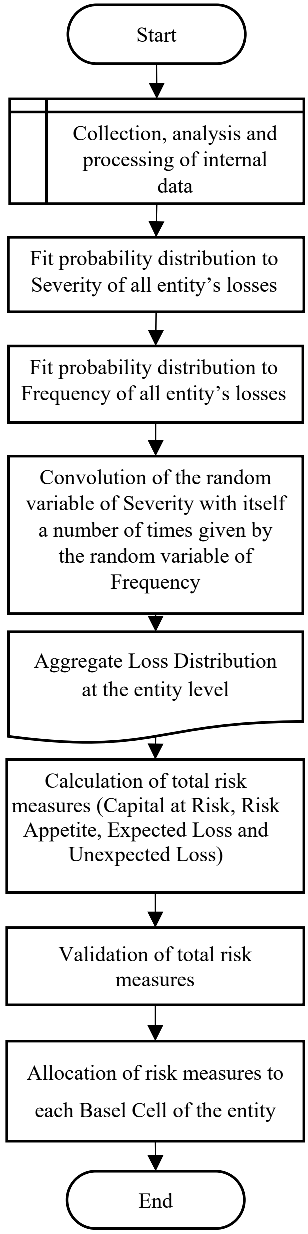

Figure 2. Flow chart for obtaining the risk measures (Capital at Risk, Risk Appetite, Expected Loss and Unexpected Loss). Own elaboration

3. Methodology

In this section, the methodology is explained, which follows the steps in *Figure 1*. For more detail, the methodology is widely described in the Master's Thesis of González (2020).

3.1. Collection, analysis and processing of internal loss data

The process begins with the collection and processing of the entity's historical internal loss data. This constitutes an essential activity to achieve the development and functioning of a credible operational risk measurement system since this makes it possible to link the risk estimates with its actual loss experience.

Internal loss data is the most relevant entry, as it is clearly linked to the bank's current business activities, technology processes, and risk management procedures. They allow to reflect the risk profile of the entity. Therefore, external data and scenarios have not been considered (see *Table 1* for the justification).

For this reason, all data must be analysed and processed with the following associated information: business line, risk type, country, currency, occurrence date, detection date, accounting date, gross loss, recovery, provision and net loss.

The date used is the first accounting date. The use of the first accounting date, instead of the date of occurrence or the detection date, is considered more appropriate because it allows for better reconciliation and validation and is consistent with the principle of prudence.

The type of loss used is the net loss, defined as the loss after all recoveries, except insurance. The use of net losses, rather than gross losses, is considered more appropriate because they collect relevant information regarding the operational risk framework, and furthermore, because gross losses would result in an artificially high Capital endowment.

Under Basel II, internally generated operational risk measures used for Economic Capital and Risk Appetite purposes should be based on a minimum observation period of five years of the most recent internal loss data and a maximum of 10 years is recommended. The time window of the data used for the proposed model is 8.5 years for greater reliability, comprising from 1st of January of 2011 to 30th of June of 2019.

To make the losses comparable during the period of time considered, it is necessary to make an inflation adjustment based on the Consumer Price Index (CPI). The inflation-adjusted losses are obtained using the following formula in *(3.1)*:

$$loss^{*}_{mm/yyyy} = loss_{mm/aaaa} \frac{CPI_{mm/yyyy^{*}}}{CPI_{mm/yyyy}} \tag{3.1}$$

where

- $loss_{mm/yyyy}$: original loss accounted for on the date mm / yyyy;
- $CPI_{mm/yyyy}$: CPI of the corresponding date;
- $CPI_{mm/yyyy^{*}}$: reference value of the IPC index, mm / yyyy * is set as 06/2019.

3.2. Severity modelling

Once the data is processed, the next step is the modelling of Severity.

The Severity Distribution of all loss data is configured, fitting various distributions and checking their goodness of fit to choose the best one. If none provide a good fit, a threshold would be imposed and various distributions would be fitted to the body and tail, checking their goodness of fit to choose the best one. If the goodness of fit fails again, mixtures will be chosen and their goodness of fit will be checked to choose the best one. If it failed again, the empirical distribution would be fitted finally.

In the following subsections, the Severity Distribution is defined, the calibration process is explained, the principles for imposing thresholds are established and the Goodness of Fit Tests are detailed.

Definition of the Severity Distribution

Severity is defined as the monetary amount of the loss associated with a loss event.

The probability distribution function of Severity is expressed as:

$$F(x) = P(\boldsymbol{X} \leq x) \tag{3.2}$$

where $\boldsymbol{X}$ is a random variable that represents the Severity.

The Severity Distribution is continuous, which can range from a simple distribution to mixtures of two components. The choice of it will be based on the one that best fits the observed historical data and best describes the underlying pattern of the loss data, emphasizing that the fit in the tail is appropriate, since this is what the Capital at Risk and Risk Appetite estimates are based, and this is why, when necessary, the adjusted distribution could be divided into two parts, the body and the tail.

Calibration of the Severity distribution

For the calibration of the Severity Distribution, that is, the estimation of its parameters, the Maximum Likelihood method is used, relying on authors such as Tan and Chang (1972), Holgersson and Jorner (1978) and Day (1969). These authors defend the Maximum Likelihood method against other point estimation methods such as the Moments Method, Bayesian Estimators, Least Squares, the Minimum Chi-Square estimation and graphic procedures.

All events are assumed to be independent random data from the same distribution.

Principles for imposing thresholds

Principles for imposing thresholds are defined here. Thresholds will be needed in case it is not possible to fit a single distribution since the use of them introduces an additional parameter to the fit. In such a case, the choice of it is very important. In this sense, it is convenient to distinguish between the body threshold and the tail threshold.

i. Body threshold. It is defined by Ferreras Salagre (2008) as “the minimum amount of the operational event from which the entity begins to collect or capture its data”. It may be applied to those losses that do not have a material impact on the Capital calculations. If applied, the distribution function will be conditioned by the mentioned threshold.

ii. Tail threshold. It is defined by Ferreras Salagre (2008) as “the threshold that the entity uses to model its data and which, logically, it must be equal to or greater than that of the body”. It divides the losses in the body and tail of the distribution. To determine the threshold, the decision is based on imposing that threshold that maximizes the number of events in the tail. The rationale behind this criterion is to cover as much data as possible to provide more information for Goodness of Fit tests.

Goodness of Fit tests for Severity Distribution

Goodness-of-Fit tests determine if the fit of the selected distribution is appropriate for the data. Refer for more detail of the Goodness of Fit tests to the Master's Thesis of González (2020).

The tests used for Severity, as it is continuous, are the *Kolmogorov-Smirnov (KS)* and *Anderson-Darling (AD)* tests. The hypotheses for these tests are:

- H_0: Data follow specified distribution;
- H_1: Data does not follow specified distribution.

The result of each one of these Goodness of Fit tests is a p-value and the significance level chosen a priori is $\alpha = 0.01$. A p-value greater than or equal to 0.01 will allow not rejecting the null hypothesis that the data follow the specified distribution and a p-value less than 0.01 will reject the null hypothesis and conclude that the data does not come from the distribution in question.

These two tests are considered conclusive for choosing the model, however Anderson-Darling Test is considered more robust because the sample size used is large and because the Capital at Risk and Risk Appetite estimates are based on the tail values.

In addition to these tests, the information criterions AIC and BIC are also calculated to provide some more information. The preferable model will be the one with the lowest AIC and BIC values.

Finally, it must be verified that the probability distribution selected in the KS and AD provides a good fit in the tail. This is a critical point because, as mentioned, the estimates of Capital at Risk and Risk Appetite are based on the tail values due to the use of the Value at Risk metric. This in reinforced by Carrillo and Suárez (2006), who argue that more than 90% of the Capital is due to a very reduced number of events that occur in the tail of the distribution.

3.3. Frequency modelling

Once the Severity Distribution is configured, the Frequency Distribution of all the loss data is proceeded to be configured. Again, various distributions would be fitted and their goodness of fit would be checked to choose the best one. If any of them provide a good fit, the empirical distribution would be fitted.

In the following subsections, the Frequency Distribution is defined, the calibration process is explained and the Goodness of Fit Test is detailed.

Definition of Frequency Distribution

Frequency is defined as the total number of operational risk events that lead to losses in a given period of time. Using this number of incidents, we fit a distribution to model the Frequencies. Thus, the Frequency Distribution describes the probability that a certain number of events will occur within a certain period of time. Note that the Frequency Distribution is discrete.

Since the Economic Capital is calculated on the basis of a time horizon of one year, the Frequency Distribution must represent the number of losses in a year. However, fitting a distribution for the random variable of Frequency requires a reasonably large amount of historical data, more than is available for this model. This drawback can be avoided by assuming that the frequencies of weekly loss are independent. These weekly frequencies can be used to fit the distributions, thus expanding the amount of data available and allowing for reliable fits. Therefore, the proposed model will use weekly frequency.

The loss Frequency Distribution function is expressed as:

$$P(n) = P(\boldsymbol{N} = n) = \sum_{k=0}^{n} p\ (k) \qquad (3.3)$$

where $\boldsymbol{N}$ is a random variable that represents the number of events that have occurred in a time horizon, whose probability function is p.

Calibration of Frequency Distribution

The Frequency Distribution is calibrated by the Maximum Likelihood method.

In the event that thresholds have been imposed on the Severity Distribution, the Frequency Distribution must be scaled taking into account the period of the internal data and the thresholds imposed on Severity using the following factor:

$$f = \frac{365}{d} \frac{\sum_{i=1}^{n} \frac{w_i}{\left(1 - cdf_{sev}^{left}(\alpha)\right)}}{\sum_{i=1}^{n} w_i} \tag{3.4}$$

where:

- d: Frequency aggregation time unit (in days);
- cdf_{sev}^{left}: limit to the left of the cumulative density function of the Severity Distribution;
- $\{w_i\}$ ni = 1: internal event weights for Severity adjustment;
- α: body fit threshold referring to internal events used for Severity fit.

Goodness of Fit test for Frequency Distribution

The Goodness of Fit test used for Frequency, as it is discrete, is the *Anderson-Darling (AD)* test.

3.4. Empirical cumulative distribution function

This section describes the empirical cumulative distribution function, which will be applied to Severity and/or Frequency in the following cases:

i. when all possible distributions have already been tested, considering simple distributions, the application of thresholds and even mixtures of two components, and none provide an appropriate fit according to Goodness of Fit tests;

ii. for the allocation process of total risk measures (at entity level) to each of the Cells. This allocation is based on the proportion of their empirical risk measure with respect to the corresponding total risk measure (explained in sec. *3.8.*).

An empirical cumulative distribution function (ECDF) is a nonparametric estimator of the underlying cumulative distribution function of a random variable. It is known as a "step" function that jumps $\frac{1}{n}$ at each step in a set of n observations, assigning a probability of $\frac{1}{n}$ to each observation in the sample. Therefore, the ECDF is a discrete cumulative distribution function that creates an exact match to the distribution of the data.

Let $X_1, \ldots, X_n$ a random variable i.i.d. in $\mathbb{R}$ with d.f.c. F(x), the ECDF, denoted by $\widehat{F}_n$, is defined as:

$$\widehat{F}_n(t) = \widehat{P}_n(X \leq t) = \frac{\#\, observations \ \leq t}{n} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{x_i \leq t\}} \tag{3.5}$$

where $\mathbb{1}_{\{.\}}$ is the indicator function given by:

$$\mathbb{1}_{\{x_i \leq x\}} = \begin{cases} 1 & si\ x_i \leq x \\ 0 & si\ x_i > x \end{cases} \tag{3.6}$$

3.5. Aggregate Loss Distribution

Once a distribution has been fit to Severity and another distribution to Frequency, the Aggregate Loss Distribution $G_S(x)$ in *(3.8)* is generated.

The aggregate loss corresponding to a temporary unit (taken as a week) is calculated as the sum of random losses corresponding to a temporary unit, whose formula is in *(3.7):*

$$S = \boldsymbol{X}_1 + \boldsymbol{X}_2 + \cdots + \boldsymbol{X}_N = \sum\nolimits_{n=0}^{\boldsymbol{N}} \boldsymbol{X}_n \tag{3.7}$$

In order to obtain the Aggregate Loss Distribution, it is done the convolution of the random variable of Severity with itself a number of times given by the random variable of Frequency.

Let $G_S(x)$ the distribution function of ϑ, $G_S(x)$, is then the aggregate distribution which is obtained by:

$$G_S(x) = \begin{cases} \displaystyle\sum_{n=1}^{\infty} P(n)\, F_X^{n^*}(x) & para\ x > 0 \\ P(0) & para\ x = 0 \end{cases} \tag{3.8}$$

where

- $P(n)$ is the Frequency Distribution function, the formula of which comes in *(3.3)*;

– $F_X^{n^*}(x)$ is n-times the convolution of F with itself. And $F_X(x)$ is the density function of $\boldsymbol{X}$ that expresses the probability that the aggregate Severity of n losses is x. To better understand convolution:

$$F^{1^*} = F \tag{3.9}$$

$$F^{n^*} = F^{(n-1)}\,F \tag{3.10}$$

Emphasize that the random variables $\boldsymbol{X}$ are implicitly assumed to be independently distributed and independent of the number of events.

Figure 3 shows a representation of the estimation process of the Aggregate Loss Distribution.

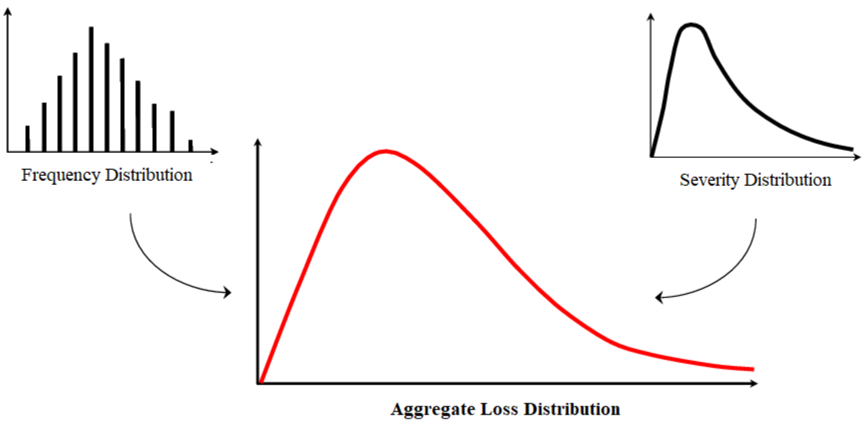

Figure 3. Representation of the estimation process of the Aggregate Loss Distribution. Own elaboration

There is no analytical expression to obtain the Aggregate Loss Distribution $G_S(x)$ in *(3.8)* and it is necessary to apply numerical algorithms such as the Fast Fourier Transform, the Panjer recursion or the Monte Carlo simulation. This Paper proposes the Monte Carlo simulation.

Next, the algorithms according to the different typology of the Severity Distribution are indicated:

Monte Carlo simulation for the case of a single distribution fitted to Severity

Let *M = {365 days, 52 weeks}* simulations of Monte Carlo where each simulation represents a simulated time unit within a year (note that *M = 52* weeks has been chosen in this Thesis). And let *C = 1,000,000* Monte Carlo simulations where each simulation represents a simulated year.

1. *Generate a random data,* ***N****, from the Frequency Distribution that determines the number of risk events that are predicted to occur in a week.*
2. *Generate* ***N*** *random values of the Severity Distribution forming a sample of simulated Severities:* $\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_N$.
3. *Sum the simulated Severities from step 2 to get the value of Aggregate Loss in a simulated time unit i:* $s_i = \boldsymbol{X}_1 + \boldsymbol{X}_2 + \cdots + \boldsymbol{X}_N$.
4. *Repeat steps 1, 2 and 3 M times to obtain a sample of Aggregate Losses in one year:* $s_1, s_2, \dots, s_M$.
5. *Sum the losses from step 4 to get the value of the Aggregate Loss in a simulated year i:* $S_i = s_1 + s_2 + \cdots + s_M$.
6. *Repeat steps 4 and 5 C times to obtain the vector of C Aggregate Losses that forms the Aggregate Loss Distribution* $G_S(x)$*:* $S_1, S_2, \dots, S_C$.

Monte Carlo simulation for the case of a two-component mixture fitted to Severity

Let be a mixture formed by two components, "component one" and "component two", where component one has a weight *w* and component two has a weight *1-w*. Let be *M = {365 days, 52 weeks}* Monte Carlo simulations where each simulation represents a time unit within a year (note that *M = 52* weeks has been chosen in this Thesis). Let be *C = 1,000,000* Monte Carlo simulations where each simulation represents a simulated year.

1. *Generate a random data,* ***N****, from the Frequency Distribution that determines the number of risk events that are predicted to occur in a week.*
2. *Generate* ***N*** *random values from a Bernoulli distribution with parameter w (the weight of component one of the mixture), N ~ Be (w), which will take value 0 with probability 1-w and value 1 with probability w.*
3. *Of those N values, generate the number of values that are 1 of the Severity Distribution of component one of the mixture, forming a sample of simulated losses:* $\boldsymbol{X}_1, \boldsymbol{X}_2, \dots$.
4. *Of those N values, generate the number of values that are 0 of the Severity Distribution of component two of the mixture, forming a sample of simulated losses:* $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots$.

5. *Sum the simulated Severities from steps 3 and 4 to get the value of the Aggregate Loss in a simulated time unit i:* $s_i = \boldsymbol{X}_1 + \boldsymbol{x}_1 + \boldsymbol{X}_2 + \boldsymbol{x}_2 + \cdots$.
6. *Repeat steps 1, 2, 3, 4 and 5 M times to obtain a sample of Aggregate Losses for one year:* $s_1, s_2, \dots, s_M$.
7. *Sum the losses from step 6 to get the value of the Aggregate Loss in a simulated year i:* $S_i = s_1 + s_2 + \cdots + s_M$
8. *Repeat steps 6 and 7 C times to obtain the vector of C Aggregate Losses that forms the Aggregate Loss Distribution* $G_S(x)$*:* $S_1, S_2, \dots, S_C$.

Monte Carlo simulation for the case of a Severity Distribution with a tail threshold imposed

Let be a Severity Distribution with a tail threshold that separates it into a body and a tail, where the body has one distribution fitted and the tail has another distribution fitted. Let's call *p* the top cumulative probability where the tail begins. Let be *M = {365 days, 52 weeks}* Monte Carlo simulations where each simulation represents a simulated time unit within a year (note that *M = 52* weeks has been chosen in this Thesis). And let be *C = 1,000,000* Monte Carlo simulations where each simulation represents a simulated year.

1. *Generate a random data **N** from the Frequency Distribution that determines the number of risk events that are predicted to occur in a week.*
2. [2]*Generate **N** random values of the Uniform distribution (0, 1).*
3. *Of these **N** values, generate the number of values less than p of the distribution fitted to the body of the Severity Distribution, forming a sample of simulated losses:* $\boldsymbol{X}_1, \boldsymbol{X}_2, \dots$.
4. *Of these **N** values, generate the number of values greater than or equal to p of the distribution fitted to the tail of the Severity Distribution, forming a sample of simulated losses:* $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots$.

[2] Steps 2, 3 and 4 could also be done by applying the Bernoulli distribution (p), where *p* indicates the highest cumulative probability where the tail begins. Generate ***N*** random values from a Bernoulli distribution with parameter *p*, *N ~ Be (p)*, which will take value *0* with probability *1-p* and value *1* with probability *p*. Of those ***N*** values, generate the number of values that are *1* of the distribution fitted to the body of the Severity Distribution and generate the number of values that are *0* of the distribution fitted to the tail of the Severity Distribution.

5. *Sum the simulated Severities from steps 3 and 4 to get the value of the Aggregate Loss in a simulated time unit i.* $s_i = \boldsymbol{X}_1 + \boldsymbol{x}_1 + \boldsymbol{X}_2 + \boldsymbol{x}_2 + \cdots$.
6. *Repeat steps 1, 2, 3, 4 and 5 M times to obtain a sample of Aggregate Losses in one year:*$s_1, s_2, \dots, s_M$.
7. *Sum the losses from step 6 to get the value of the Aggregate Loss in a simulated year i:* $S_i = s_1 + s_2 + \cdots + s_M$.
8. *Repeat steps 6 and 7 C times to obtain the vector of C Aggregate Losses that forms the Aggregate Loss Distribution* $G_S(x)$*:* $S_1, S_2, \dots, S_C$.

After completing the Monte Carlo simulation process, the data of the resulting vector are ordered from lowest to highest and the 99.90th and 95th percentiles that determine the Capital at Risk and Risk Appetite, respectively, are calculated.

The following section, 3.6, details the mentioned risk measures.

3.6. Risk measures

The risk measures, which use the Aggregate Loss Distribution as input, are the following:

Capital at Risk (CaR)

It represents the amount of Capital necessary to cover the Expected and Unexpected Losses due to operational risk that may be originated in a time horizon of one year with a confidence level of 1- α. It is based on the Value at Risk metric, and more specifically on the Operational VaR (OpVaR) metric. Its formula is in *(3.11)*:

$$CaR(1-\alpha) = OpVar(1-\alpha) = G_S^{-1}(1-\alpha) \tag{3.11}$$

Frachot, Georges & Roncalli (2001) suggest that the percentile (1-α) for purposes of Economic Capital should be established based on the Entity's rating. *Table 3* shows the values of α:

Table 3. Values of the percentiles according to the rating

Rating	BBB	A	AA	AAA
1-α	99,75%	99,90%	99,95%	99,97%

Source: Frachot et al. (2001)

Assuming that the Entity for which its capital will be calculated has an A rating, according to Standard & Pool's, Moodys and Fith, the (1-α) used is 99.90%. Therefore, Capital at Risk, in this Paper, will be defined as the 99.90th percentile of the vector of one million Aggregate Losses that forms the Entity's Aggregate Loss Distribution, for a time horizon of one year.

Expected Loss (EL)

It represents the potential loss that under normal conditions the entity expects to suffer in a period of one year. Therefore, it estimates loss values with a greater probability of occurrence. This risk measure is defined as the mean of the vector of one million Aggregate Losses that forms the Aggregate Loss Distribution.

Unexpected Loss (UL)

It represents events not foreseen by the entity with less probability of occurrence, but with very high loss values that would significantly impact the Capital of the financial entity. This risk measure is defined as the difference between the Capital at Risk and the Expected Loss (eq. *(3.13)*).

$$CaR = EL + UL \quad (3.11)$$
$$UL = CaR - EL \quad (3.12)$$

Risk Appetite (RA)

It represents the amount of risk that an entity is willing to seek and assume in pursuit of its objectives. This risk measure is defined as the 95th percentile of the vector of one million Aggregate Losses that forms the Aggregate Loss Distribution.

Figure 4 shows a graph of the risk measures indicated in the Aggregate Loss Distribution:

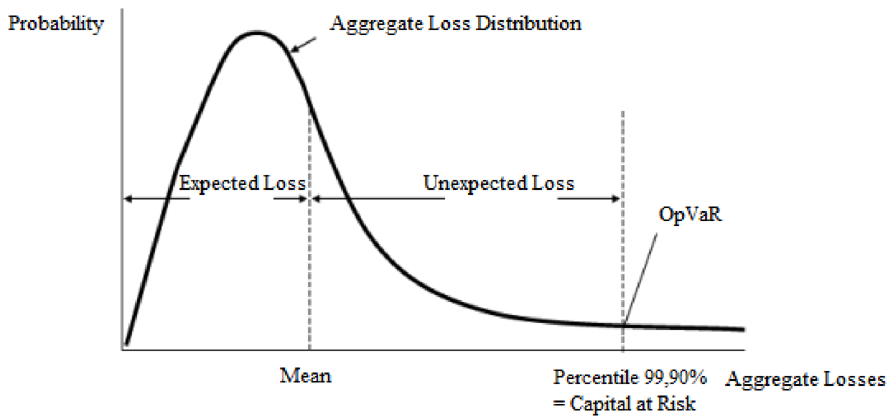

Figure 4. Representation of risk measures in the Distribution of Aggregate Losses. Own elaboration

Once the risk measures are obtained, these should be validated, which is explained in the following section.

3.7. Validation of total risk measures

In case parametric distributions to Severity and/or Frequency could be fitted, the validation process consists on verifying that the "parametric total risk measures" are superior to the "empirical total risk measures", where parametric risk measures refer to those obtained by the convolution of the continuous parametric random variable of Severity with itself a number of times given by the discrete parametric random variable of Frequency and empirical risk measures refer to those obtained by the convolution of the continuous empirical random variable of Severity with itself a number of times given by the discrete empirical random variable of Frequency.

The justification behind this methodology is because the empirical distribution is based only on realized values and is not conservative. In addition, it should be verified that the figures of these measures are reasonable.

3.8. Allocation of total risk measures to each of the Cells of Basel

The model uses a top-down approach, which consists of the the calculation of the total risk measures (at entity level) in the first instance, which has been already explained, and afterwards the disaggregation of these measures among the Basel Cells presented by the Entity, which is going to be explained in this section.

Each cell is a combination between a BL and a RT. This categorization is based on the principle of organizing losses into categories that share the same basic risk profile and behavior pattern, forming an array with 56 data Cells, showed in *Figure 1*, and BLs and RTs are indicated in *Table 2*.

The allocation process is as follows: first, for each Cell, perform the convolution of the empirical r.v of Severity with itself a number of times given by the empirical random variable of Frequency (see sec. *3.4.*) using the method of Monte Carlo Simulation to obtain the empirical vector of one million Aggregate Losses corresponding to each Cell (see sec. *5.4* - *Monte Carlo simulation for the case of a single distribution fitted to Severity*. Next, for each Cell, calculate the empirical risk measures (see sec. *3.6.* and subsequently calculate the proportion that each represents with respect to its total empirical risk measure. The risk measures assignable to each Cell will be the result of multiplying this proportion by the corresponding entity's risk measure.

To specify more, the risk measures assignable to Cell i are calculated with the following formulas:

$$CaR_{\mathrm{i}} = \frac{Empricial\ CaR_{\mathrm{i}}}{\sum_{i=1}^{56} Empirical\ CaR_i} CaR_{entity}, \qquad i = 1, \dots, 56 \tag{3.13}$$

$$RA_{\mathrm{i}} = \frac{Empirical\ RA_{\mathrm{i}}}{\sum_{i=1}^{56} Empirical\ RA_{\mathrm{i}}} RA_{entity}, \qquad i = 1, \dots, 56 \tag{3.14}$$

$$EL_{\mathrm{i}} = \frac{Empricial\ EL_{\mathrm{i}}}{\sum_{i=1}^{56} Empricial\ EL_{\mathrm{i}}} EL_{entity}, \qquad i = 1, \dots, 56 \tag{3.15}$$

$$UL_{\mathrm{i}} = \frac{Empirical\ UL_{\mathrm{i}}}{\sum_{i=1}^{56} Empirical\ UL_{\mathrm{i}}} UL_{entity}, \qquad i = 1, \dots, 56 \tag{3.16}$$

where

- $Empirical\ CaR_i, Empirical\ RA_i, Empirical\ EL_i, Empirical\ UL_i$: risk measures corresponding to Cell i obtained using as an input the empirical vector of one million Aggregate Losses of Cell i. Note that distributions of Severity and Frequency would be empirical;
- $CaR_{entity}, RA_{entity}, EL_{entity}, UL_{entity}$: total risk measures attributable to the Entity obtained using as an input the total vector of one million Aggregate Losses (at entity level).

The reason behind the methodological proposal presented which consists of a top-down approach, rather than a bottom-up approach which consists of the calculation of the total risk measures (at the entity level) as the sum of the risk measures of each Risk Unit, lies in the existence of two problems, one is that numerous Cells have insufficient sample size and the other one is that the shape of the distribution makes modelling impossible, having to resort to the use of the empirical distribution or in distributions with many parameters giving rise to overfitting problems. This reasoning is justified by the analysis carried out in section *4*.

The benefit of this methodology is twofold: i) it solves the problem of data scarcity since the sample is large enough to be modelled with a parametric distribution of few parameters, and ii) it is not necessary to resort to external data to enlarge the sample, since external data have a completely different nature as it comes from other financial entities (see *Table 1* for more information), managing to provide robust and reliable estimates.

4. Data

The sample of data used for the development and implementation of the proposed model is based solely on historical internal loss data due to operational risk generated from plausible loss data for a credit institution. Hereinafter, this anonymous entity will be referred to by the name "Entidad SA". For confidentiality reasons, the data has been slightly modified, which may prevent the real comparison with the market, but does not invalidate the proposed methodology.

The database contains the following variables: business line, risk type, country, currency, occurrence date, detection date, accounting date, gross loss, recovery, provision and net loss.

The purpose of the present section is to perform a study of modelling of the Basel Cells and Business Lines in order to reach the conclusion of whether to model at the Cell level, at the Risk Unit level or at the Entity level.

According to the methodology proposed by Basel II, it recommends estimating the distribution of losses individually for each of the Cells (Comité de Supervisión Bancaria de Basilea, 2006), which are made up of a Business Line and a Risk Type. This would be the ideal situation and could be carried out if each of the Cells had enough data and an appropriate shape of the distribution to be modelled, so that both Severity and Frequency data could be modelled without falling into overfitting problems or resorting to empirical distribution. Next, we will see that this methodology cannot be carried out due to the aforementioned problem.

For this, two studies have been performed: i) the analysis of the Severity and Frequency Distributions of each of the Basel Cells of which the entity is composed, and ii) the analysis of the Severity and Frequency Distributions of the Business Lines. The pursuit of both studies is the analysis of the sample size and the shape of the distributions of Frequency and Severity.

Study of modelling of Cells

Figure 5 shows the Severity histograms of each of the Cells whose ordinate axis indicates the number of losses and the abscissa axis indicates the Severity in logarithmic scale of base 10 and *Figure 6* shows the Frequency histograms of each of the Cells whose ordinate axis indicates the probability and the abscissa axis indicates the number of events per week.

In both Figures, it can be seen that most of the Cells could not be modelled due to two problems, the insufficient sample size that they present and the shape of the distribution that makes modelling impossible having to resort to the use of the empirical distribution which is considered to be not very conservative since it only takes into account realized values or having to fit distributions with many parameters giving rise to overfitting problems.

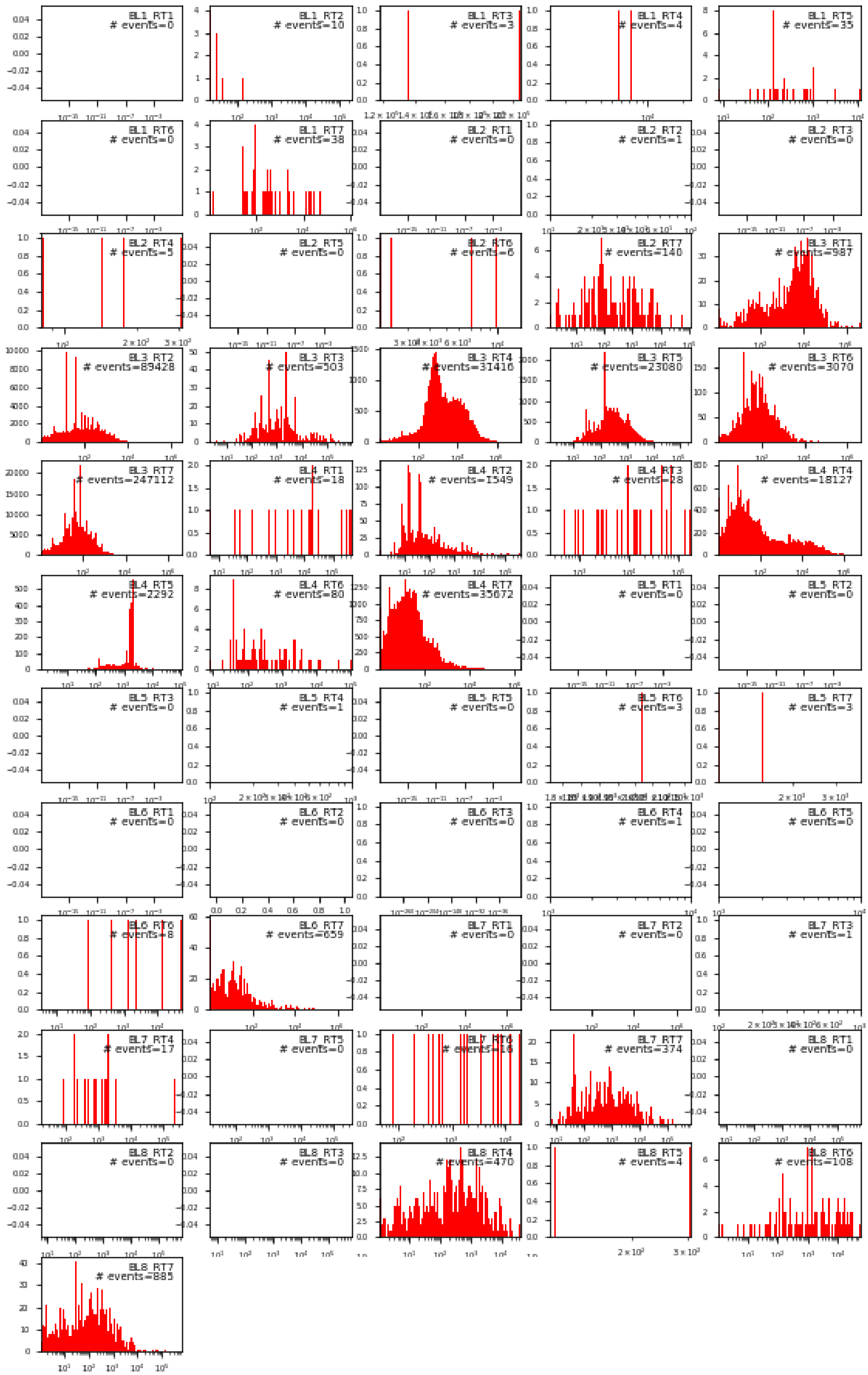

Figure 5. Severity histograms of the Basel Cells. Source: Own elaboration

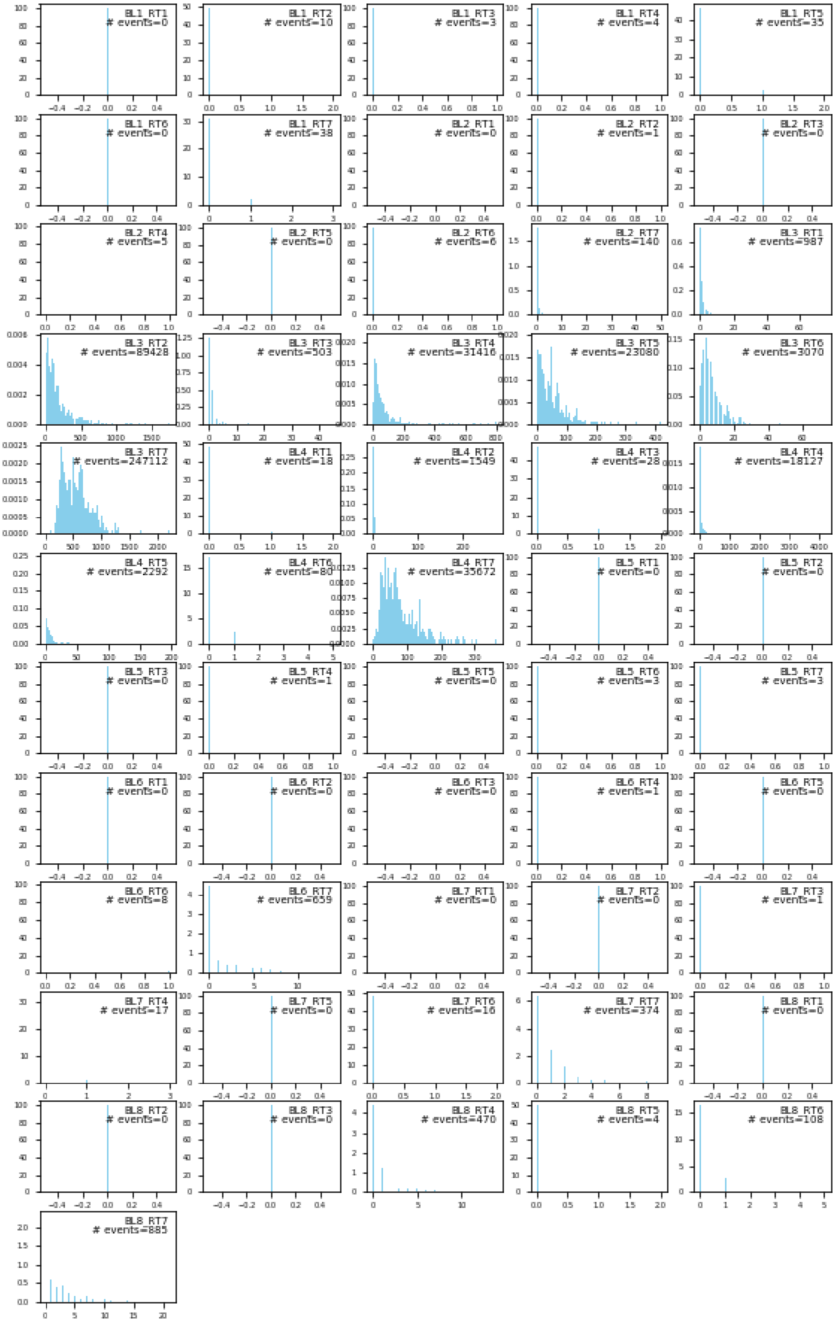

Figure 6. Frequency histograms of Basel Cells. Source: Own elaboration

Study of modelling of Business Lines

In this context, as the option of modelling at Cell level has been discarded, the need arises to make Cell aggregations to form Risk Units. The criterion for forming Risk Units will be to take the Business Line as axis.

Figure 7 shows the Severity histograms of each of the Business Lines whose ordinate axis indicates the probability and the abscissa axis indicates the Severity in logarithmic scale of base 10 and *Figure 8* shows the Frequency histograms of each of the Business Lines whose ordinate axis indicates the probability and the abscissa axis indicates the number of events per week.

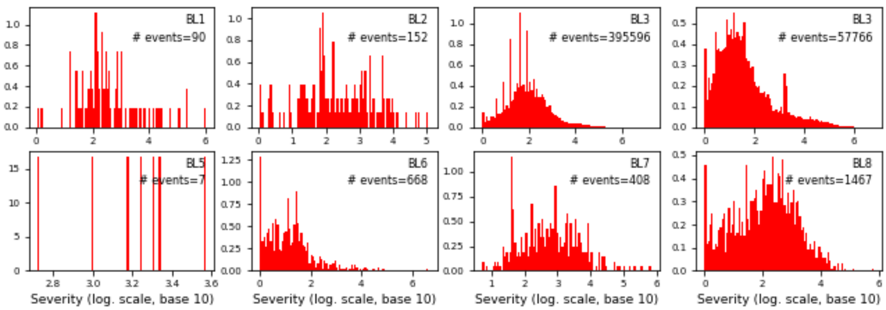

Figure 7. Severity histograms of the Business Lines. Source: Own elaboration

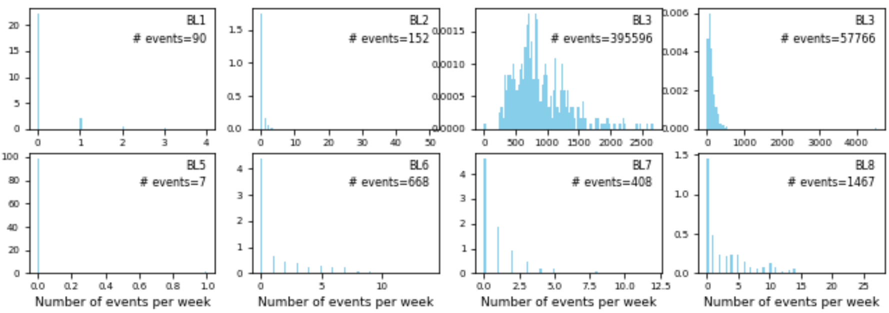

Figure 8. Frequency histograms of the Business Lines

Again, it can be seen in *Figure 7* and *Figure 8 that* both problems already commented in the study of the Cells happen again, which are the insufficient sample size and the complicate shape of the distribution to be modelled. In particular, 5 of a total of 8 Business Lines have an insufficient sample size since they do not even reach the number of 1,000 events and the same number have a Frequency Distribution with a complicate to be modelled. Therefore, modelling at Risk Unit level is discarded.

After both studies, it is concluded that a segmentation of the data a priori to form Cells or Risk Units is not consistent because it makes the size of some samples insufficient to be able to model them. Therefore, the solution implemented in this model follows a top-down approach which consists of estimating a joint Aggregate Loss Distribution that encompasses all Cells (at entity level), calculating the risk measures (Capital at Risk, Risk Appetite, Expected Loss and Unexpected Loss) that would be the total risk measures for the entity and afterwards the disaggregation of them for each Cell.

The benefit of this methodology is twofold: i) it solves the problem of data scarcity since the sample is large enough to be modelled with a parametric distribution of few parameters, and ii) it is not necessary to resort to external data to enlarge the sample, since external data have a completely different nature as it comes from other financial entities (see *Table 1* for more information), managing to provide robust and reliable estimates.

5. Model implementation

The implementation of the model takes place in this section. To do this, the steps specified in section *2* will be followed.

5.1. Collection, analysis and processing of data

The first step is the collection, processing and inflation adjustment of the data. This process is detailed in section *3.1.*

5.2. Severity modelling

The second step is the configuration of the Severity Distribution, the foundations of which are found in section *3.2.* For this, the log-normal, log-logistic, burr and beta prime distributions and mixture of two log-normals have been fitted to Severity data and all of them have subsequently been subjected to goodness-of-fit tests to determine which one provides the best fit.

Calibration of probability distributions

The parameters of the mentioned probability distributions have been estimated by the Maximum Likelihood method. They are shown in *Table 4.*

Table 4. Parameters of probability distributions fitted to the Severity data

	Shape parameters	**Location parameters**	**Scale parameters**	**Weights (only mixtures)**
Log-normal[3]	s=1.96156	loc=0.49120	scale=756.51351	
Log-logistic (fisk)[4]	c=0.92558	loc=1.00100	scale=682.30838	
Burr[5]	c=0.78208 d= 2.00265	loc=-0.5707	scale=210.48392	
Beta Prime[6]	a= 1.30236 b= 0.71158	loc=0.68933	scale=271.37123	
Mixture of two log-normals		loc_1 = 2.63587 loc_2 = 3.09189	$scale_1$ = 0.43235 $scale_2$ = 1.01506	$weight_1$ = 0.427 $weight_2$ = 0.573

Source: Own elaboration

To offer a visual analysis of the degree of fit of each of the calibrated probability distributions, *Figure 9* and *Figure 10* show the probability density functions of each of them plotted on the empirical Severity Distribution.

[3] Another parameterization of the log-normal distribution: σ=s, μ=ln(scale).
[4] Another parameterization of the log-logistic distribution: β=c, α=scale.
[5] Another parameterization of the burr distribution in Scipy: c= c, k=d.
[6] Another parameterization of the beta prime distribution: α=a, β=b.

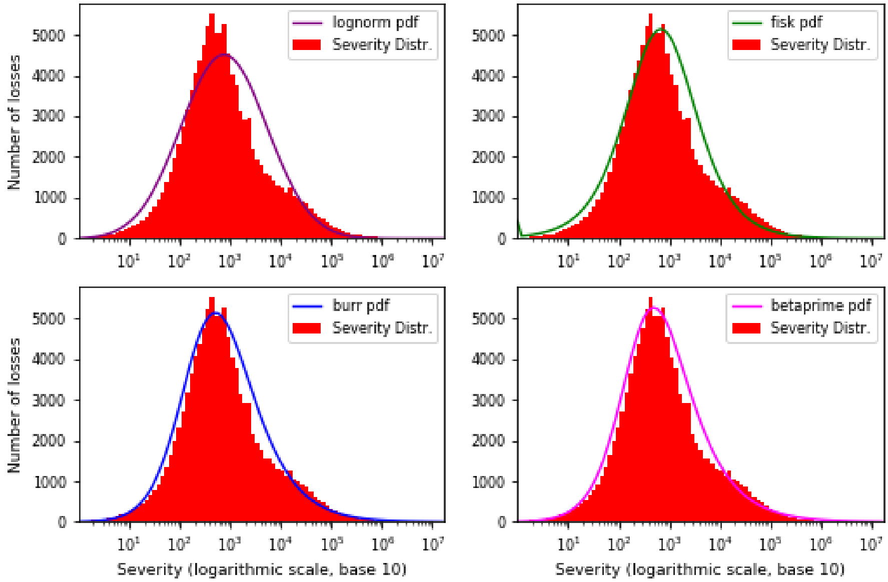

Figure 9. Probability distributions fitted to the Severity data. Own elaboration[7]

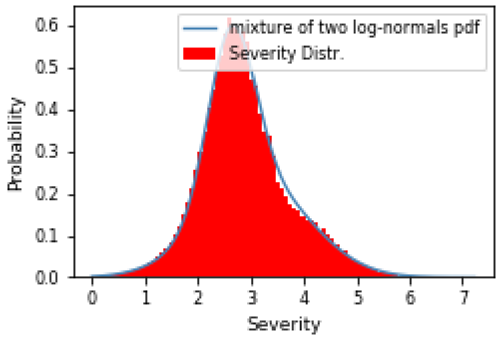

Figure 10. Mixture of two log-normals fitted to Severity data. Own elaboration[8]

---

[7] The representation of the abscissa axis is performed on a logarithmic scale of base 10 and the values indicated on the abscissa axis are the original values.

[8] The mixture of two log-normals corresponds to fitting a mixture of two normals on the logarithm of base 10 of the data.

Visually, it can be appreciated that the best-fitting probability distribution is the mixture of two log-normals (see *Figure 10*).

Goodness of Fit tests

The probability distributions have been subjected to the *Kolmogorov-Smirnov (KS)* and the *Anderson-Darling (AD)* tests. Additionally, the information criteria (AIC and BIC) have been calculated. Finally, for greater precision, the fit in the tail of the probability distribution with the highest p-value has been evaluated.

Table 5 shows the statistic and the p-value of both Goodness of Fit tests for each probability distribution.

Table 5. Statistics and P-values of the Kolmogorov-Smirnov and Anderson-Darling tests of probability distributions fitted to Severity

	Kolmogorov Smirnov		**Anderson-Darling**	
	Statistic	**P-value**	**Statistic**	**P-value** [9]
Log-normal	0.066	2.471e-35	88.727	0.001
Log-logistic (fisk)	0.0387	2.291e-12	37.119	0.001
Burr	0.026	1.343e-05	11.711	0.001
Beta Prime	0.025	3.165e-05	8.641	0.001
Mixtura de dos log-normales	**0.010**	**0.340**	**-0.119**	**0.25**

Source: Own elaboration

The conclusion drawn is that the data follow a mixture of two log-normals with a confidence level of 99%, since it is the only probability distribution with a p-value greater than or equal to $\alpha = 0.01$.

It should be noted that it is very difficult to model Severity with a single simple probability distribution function because the typology of the losses due to operational risk is very complex (Fernández-Laviada, 2010). As is evident, none of the simple probability distributions fits, so a mixture of two log-normals had to be calibrated, which does provide a satisfactory fit.

Together with the two previous tests, the AIC and BIC values present in *Table 6* have been calculated. Note that this criterion will not be conclusive for choosing the best probability distribution. According to the information

[9] Los p-valores de la prueba Anderson-Darling Two Samples tienen un límite inferior de 0,1% y un límite superior de 25%.

criteria, the preferable model will be the one with the lowest AIC and BIC values.

Table 6. AIC and BIC of the probability distributions fitted to Severity data

	AIC	**BIC**
Log-normal	1.914.430,693	1.914.459,454
Log-logístic (fisk)	3.553.251,667	3.553.280,428
Burr	1.872.742,378	1.872.780,726
Beta Prime	1.872.296,509	1.872.334,857
Mixture of two log-normals	3.453.496,187	3.453.544,122

Source: Own elaboration

It is appreciated that the probability distribution with the lowest AIC and BIC values is the beta prime. The mixture of two log-normals has higher AIC and BIC values since these two criteria penalize the number of parameters. In particular, the mixture of two log-normals has 5 parameters, a number greater than the number of parameters of the other distributions.

Nonetheless, the conclusive test is the KS and AD tests, as stated in section *3.2.* , in which the best model is the mixture of two log-normals.

Additionally, for greater precision, it has been visually verified, in *Figure 11*, that the fit in the tail of the distribution resulting from the KS and AD tests is appropriate, which was the mixture of two log-normals. The first graph shows all data greater than € 1,000 and the second graph shows all data greater than € 1,000,000.

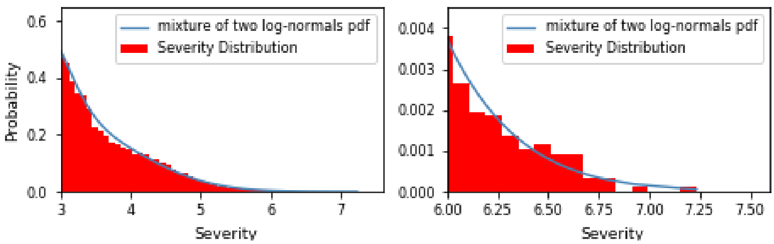

Figure 11. Fit in the tail of the mixture of two log-normals.
Source: Own elaboration

It is evident that the mixture of two log-normals provides a good fit in the tail.

Therefore, it is concluded that the probability distribution chosen to model Severity is the mixture of two log-normals.

5.3. Frequency modelling

The third step is the configuration of the Frequency Distribution, whose foundations are in section 3.3. For this, the poisson, binomial, negative binomial and geometric distributions have been fitted to Frequency data and all of them have subsequently been subjected to goodness-of-fit tests to determine which one provides the best fit.

Calibration of probability distributions

The parameters of the mentioned probability distributions have been estimated by the Maximum Likelihood method. They are shown in *Table 7.*

Table 7. Parameters of the probability distributions fitted to the Frequency data

	Shape parameters
Poisson[10]	μ = 1027,37387
Binomial[11]	n = ∄ p = ∄
Negative binomial[12]	n = 4,31144 p = 0,00418
Geometric	p = 0,00097

Source: Own elaboration

To offer a visual analysis of the degree of fit of each of the calibrated probability distributions, *Figure 12* shows the probability density functions of each of them plotted on the empirical Frequency Distribution.

[10] Another parameterization of the poisson distribution: $\lambda=\mu$.
[11] The parameters of the binomial distribution do not exist because when they were estimated they gave a negative value and according to the binomial probability theory, their parameters have the following domain: $n\geq 0$ and $0\leq p\leq 1$.
[12] Another parameterization of the negative binomial distribution: r=n, p=p.

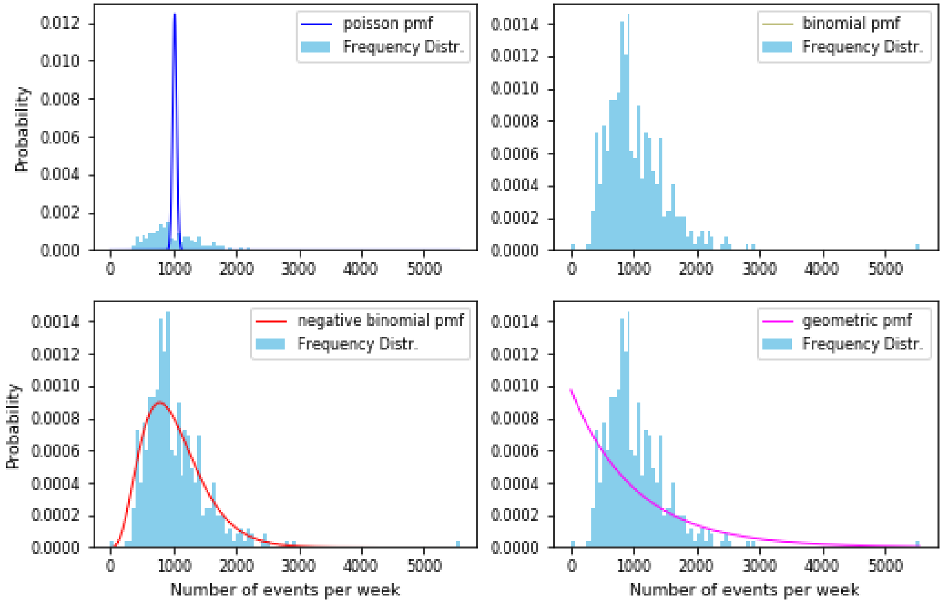

Figure 12. Probability distributions fitted to the Frequency data. Own elaboration[13]

Visually, it can be appreciated that the best-fitting probability distribution is the negative binomial distribution.

Goodness of Fit tests

The probability distributions have been subjected to the *Anderson-Darling (AD)* test since this test is feasible for discrete distributions, while the *Kolmogorov-Smirnov* test is not.

Table 8 shows the statistic and the P-value for each probability distribution.

[13] As expected, the binomial probability density function does not appear due to the fact that its parameters do not exist.

Table 8. Statistics and P-values of the Anderson-Darling test of the probability distributions fitted to Frequency

	Anderson-Darling	
	Statistic	**P-value**[14]
Poisson	137.099	0.001
Binomial	259407.994	0.001
Negative Binomial	**0.519**	**0.202**
Geometric	53.885	0.001

Source: Own elaboration

The conclusion drawn is that the data follow a negative binomial distribution with a confidence level of 99% since it is the only probability distribution with a P-value greater than or equal to $\alpha = 0.01$.

Note that it is an achievement to have been able to fit the negative binomial distribution to the Frequency data since the typology of operational risk losses is very complex and therefore it is very difficult to fit a simple probability distribution. Usually, in this context, the empirical distribution is used, which should be considered the last option since it is not considered conservative.

5.4. *Estimation of the Aggregate Loss Distribution*

Fitted a distribution to Severity and fitted another distribution to Frequency, the fifth step consists of estimating the Aggregate Loss Distribution, whose foundations are in section 3.5.

Remember that the random variable of Severity follows a mixture of two log-normals (μ_1=2.63587, σ_1=0.43235, w_1=0.427, μ_2=3.09189, σ_2=1.01506, w_2=0.573) and the random variable of Frequency follows a negative binomial distribution (n=4.31144, p=0.00418).

As already indicated, the chosen time unit is week, therefore the algorithm specified below is formulated to obtain first the Aggregate Loss of one week, then the Aggregate Loss for one year and later the Aggregate Losses of 1,000,000 years. The algorithm is as follows:

[14] The p-values of the Anderson-Darling test have a lower limit of 0.1% and an upper limit of 25%.

1. Generate a random data $\boldsymbol{N}$ from the Frequency Distribution that determines the number of risk events that are predicted to occur in a week.
2. Generate $\boldsymbol{N}$ random values from a Bernoulli distribution with parameter 0.427, $\boldsymbol{N}$ ~ Be (w), which will take value 0 with probability 0.573 and value 1 with probability 0.427.
3. Of those $\boldsymbol{N}$ values, generate the number of values that are 1 of the Severity Distribution of component one of the mixture, forming a sample of simulated losses: $\boldsymbol{X}_1, \boldsymbol{X}_2, ...$
4. Of those $\boldsymbol{N}$ values, generate the number of values 0 of the Severity Distribution of component two of the mixture, forming a sample of simulated losses: $\boldsymbol{x}_1, \boldsymbol{x}_2, ...$
5. Sum the simulated Severities from steps 3 and 4 to get the value of Aggregate Loss in a simulated week i: $s_i = \boldsymbol{X}_1 + \boldsymbol{x}_1 + \boldsymbol{X}_2 + \boldsymbol{x}_2 + \cdots$
6. Repeat steps 1, 2, 3, 4 and 5 52 times to obtain a sample of weekly Aggregate Losses in a year: $s_1, s_2, \dots, s_{52}$.
7. Sum the losses from step 6 to get the value of the Aggregate Loss in a simulated year i: $S_i = s_1 + s_2 + \cdots + s_{52}$.
8. Repeat steps 6 and 7 1,000,000 times to obtain the 1,000,000 Aggregate Loss vector that forms the Aggregate Loss Distribution $G_S(x)$: $S_1, S_2, \dots, S_{1,000,000}$.

The total Aggregate Loss Distribution (at entity level) obtained after the application of the previous algorithm is shown in *Figure 13*.

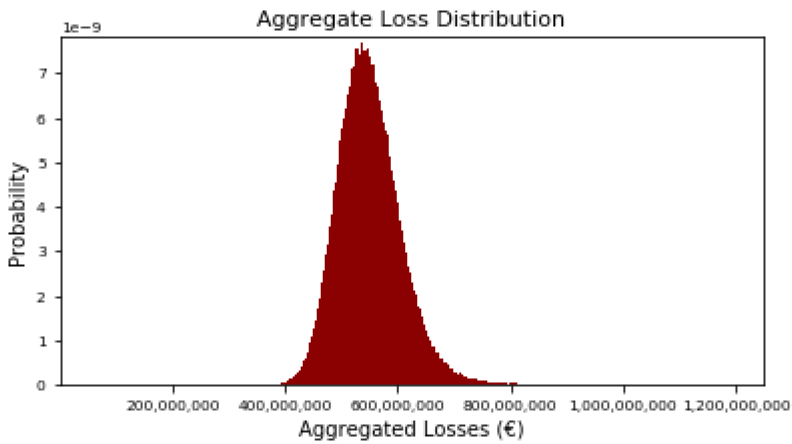

Figure 13. Aggregate Loss Distribution of the Entity SA. Own elaboration

5.5. Calculation of the total risk measures

Once the vector of Aggregate Losses has been obtained, and with it formed the Aggregate Loss Distribution at entity level, the sixth step takes place, in which the total risk measures are calculated, the foundations of which are found in section *3.6*.
The calculation of the risk measures is indicated below:

- Capital at Risk (CaR): defined as the 99.90th percentile of the vector of one million aggregate losses.
- Risk Appetite (RA): defined as the 95th percentile of the vector of one million aggregate losses.
- Expected Loss (EL): calculated as the mean of the vector of one million aggregate losses.
- Unexpected Loss (UL): calculated as the difference between the Capital at Risk and the Expected Loss.

Table 9 shows the values of the mentioned total risk measures and *Figure 14 shows* these measures represented in the entity's Aggregate Loss Distribution.

Table 9. Total risk measures (Capital at Risk, Risk Appetite, Expected Loss and Unexpected Loss) of the entity

Capital at Risk (CaR)	**Risk Appetite (RA)**	**Expected Loss (EL)**	**Unexpected Loss (UL)**
926,586,457 €	649,253,228 €	550,609,631 €	375,976,826 €

Source: Own elaboration

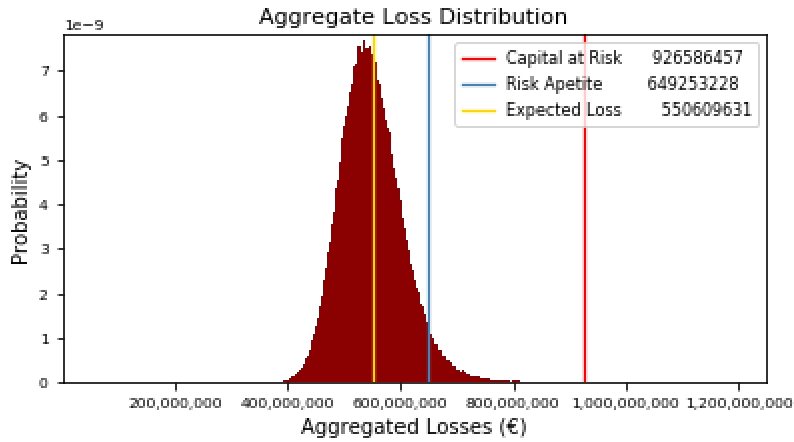

Figure 14. Representation of total risk measures in the entity's Aggregate Loss Distribution. Own elaboration

5.6. Validation of total risk measures

Once the total risk measures have been obtained, they are validated, which constitutes the seventh step, the foundations of which are found in sec. *3.7.*

The validation is based on verifying that the parametric total risk measures are superior to the empirical total risk measures. The justification behind this methodology is due to the fact that the empirical distribution is based solely on realized values, thus is not considered conservative.

Table 10 shows the figures of the total empirical and parametric total risk measures.

Table 10. Comparison of the empirical and parametric total risk measures

	Capital at Risk (CaR)	**Risk Appetite (RA)**	**Expected Loss (EL)**	**Unexpected Loss (UL)**
Empirical	694.900.233 €	614.524.080 €	539.016.216 €	155.884.017 €
Parametric	926.586.457 €	649.253.228 €	550.609.631 €	375.976.826 €

Source: Own elaboration

It is verified that the parametric total risk measures are superior to the empirical total risk measures and are also reasonable, therefore, the validation is positive.

3.9. Allocation of total risk measures to each of the Cells of Basel

Once the total risk measures (at the entity level) have been validated, the eighth step takes place, in which these total measures are disaggregated in each of the Basel Cells. The allocation process is described in section 3.8.

Figure 15 shows the Aggregate Loss Distributions for each Cell with their corresponding empirical risk measures figures (not final). The ordinate axis indicates the probability and the abscissa axis indicates the amount of aggregate losses on a logarithmic scale.

Table 11 shows the final figures of the risk measures of each Cell.

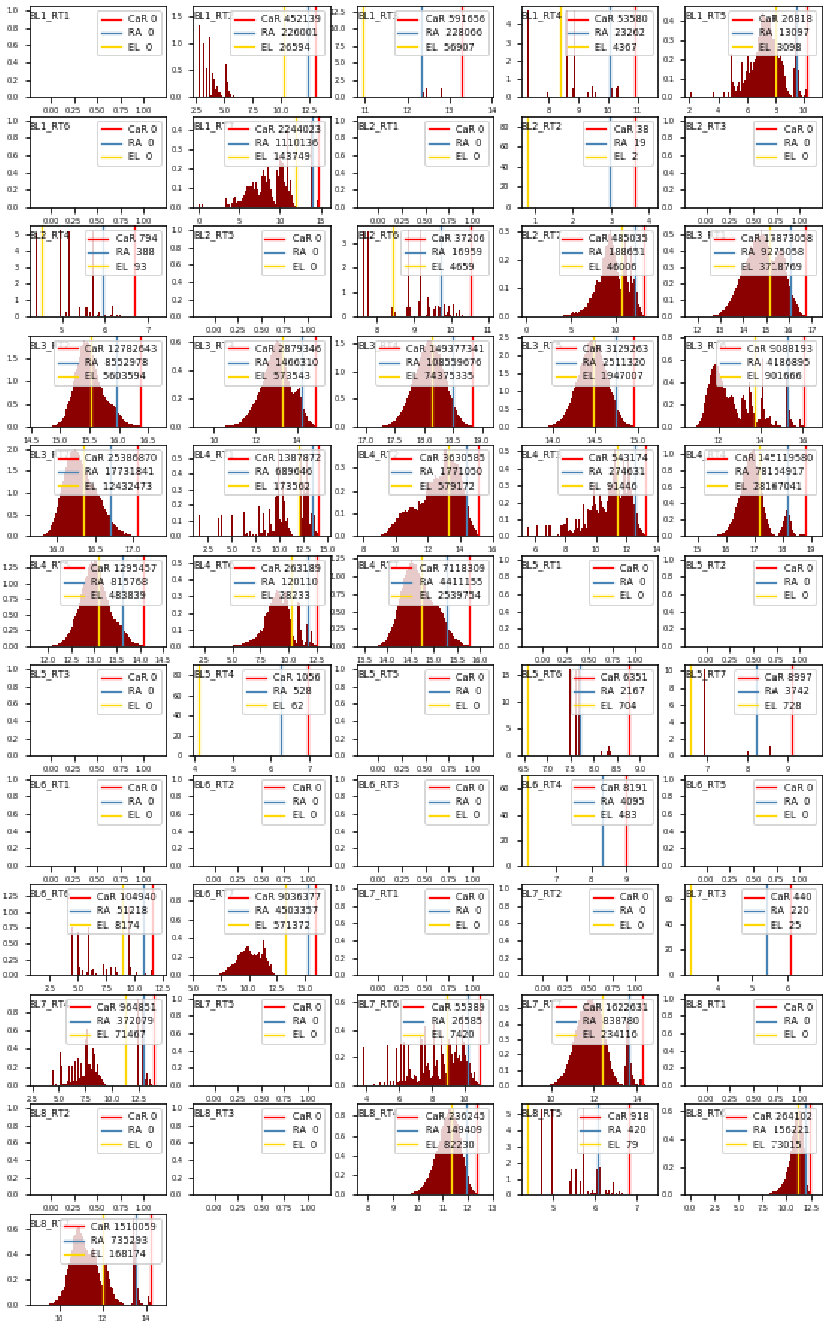

Figure 15. Aggregate Loss Distributions of the Basel Cells presented by the Entity SA. Source: Own elaboration

Table 11. Final risk measures figures (Capital at Risk, Risk Appetite, Expected Loss and Unexpected Loss) of the Basel Cells presented by the Entity SA

Basel Cells	Capital at Risk (CaR)	Risk Appetite (RA)	Expected Loss (EL)	Unexpected Loss (UL)
BL1_RT1	- €	- €	- €	- €
BL1_RT2	1.053.723 €	593.645 €	110.000 €	943.723 €
BL1_RT3	1.378.871 €	599.067 €	235.384 €	1.143.487 €
BL1_RT4	124.870 €	61.103 €	18.064 €	106.806 €
BL1_RT5	62.501 €	34.403 €	12.818 €	49.684 €
BL1_RT6	- €	- €	- €	- €
BL1_RT7	5.229.756 €	2.916.024 €	594.579 €	4.635.178 €
BL2_RT1	- €	- €	- €	- €
BL2_RT2	90 €	51 €	9 €	81 €
BL2_RT3	- €	- €	- €	- €
BL2_RT4	1.852 €	1.020 €	386 €	1.465 €
BL2_RT5	- €	- €	- €	- €
BL2_RT6	86.710 €	44.549 €	19.272 €	67.438 €
BL2_RT7	1.130.387 €	495.536 €	190.292 €	940.096 €
BL3_RT1	41.653.639 €	24.363.034 €	15.381.656 €	26.271.982 €
BL3_RT2	29.790.291 €	22.466.329 €	23.177.710 €	6.612.581 €
BL3_RT3	6.710.394 €	3.851.597 €	2.372.303 €	4.338.091 €
BL3_RT4	**348.127.867 €**	**285.156.497 €**	**307.632.900 €**	**40.494.967 €**
BL3_RT5	8.092.832 €	6.596.551 €	8.053.255 €	39.577 €
BL3_RT6	21.180.277 €	10.997.825 €	3.729.491 €	17.450.786 €
BL3_RT7	59.164.776 €	46.576.685 €	51.423.471 €	7.741.306 €
BL4_RT1	3.234.474 €	1.811.513 €	717.895 €	2.516.580 €
BL4_RT2	8.461.175 €	4.652.065 €	2.395.584 €	6.065.591 €
BL4_RT3	1.265.882 €	721.381 €	378.244 €	887.638 €
BL4_RT4	**338.205.040 €**	**205.291.535 €**	**116.505.137 €**	**221.699.903 €**
BL4_RT5	3.019.099 €	2.142.800 €	2.001.267 €	1.017.832 €
BL4_RT6	613.369 €	315.496 €	116.780 €	496.589 €
BL4_RT7	15.789.409 €	11.586.896 €	10.504.987 €	5.284.422 €
BL5_RT1	- €	- €	- €	- €
BL5_RT2	- €	- €	- €	- €
BL5_RT3	- €	- €	- €	- €
BL5_RT4	2.462 €	1.387 €	257 €	2.205 €
BL5_RT5	- €	- €	- €	- €
BL5_RT6	14.803 €	5.694 €	2.914 €	11.889 €
BL5_RT7	20.970 €	9.831 €	3.012 €	17.958 €
BL6_RT1	- €	- €	- €	- €
BL6_RT2	- €	- €	- €	- €
BL6_RT3	- €	- €	- €	- €
BL6_RT4	19.091 €	10.759 €	1.999 €	17.092 €
BL6_RT5	- €	- €	- €	- €
BL6_RT6	244.567 €	134.538 €	33.811 €	210.756 €
BL6_RT7	21.059.519 €	11.829.084 €	2.363.323 €	18.696.196 €
BL7_RT1	- €	- €	- €	- €
BL7_RT2	- €	- €	- €	- €
BL7_RT3	1.026 €	578 €	107 €	919 €
BL7_RT4	2.248.613 €	977.350 €	295.606 €	1.953.007 €
BL7_RT5	- €	- €	- €	- €

BL7_RT6	129.086 €	69.832 €	30.694 €	98.392 €
BL7_RT7	3.781.586 €	2.203.245 €	968.359 €	2.813.227 €
BL8_RT1	- €	- €	- €	- €
BL8_RT2	- €	- €	- €	- €
BL8_RT3	- €	- €	- €	- €
BL8_RT4	550.578 €	392.457 €	340.125 €	210.453 €
BL8_RT5	2.140 €	1.106 €	327 €	1.813 €
BL8_RT6	615.499 €	410.352 €	302.010 €	313.489 €
BL8_RT7	3.519.234 €	1.931.415 €	695.605 €	2.823.628 €
TOTAL	**926.586.457 €**	**649.253.229 €**	**550.609.631 €**	**375.976.826 €**

Source: Own elaboration

6. Results

6.1. Evaluation of the results of the risk measures of the Cells

This section has been designed to evaluate the results obtained in the previous section, specifically in *Table 11*, in order to identify those Cells with the highest charges for Capital at Risk, Risk Appetite, Expected Loss and Unexpected Loss, seek consistency and draw conclusions. Additionally, a recommendation is made for the budget of risk mitigation plans.

It should be noted there are two very significant Cells, which reach approximately 75% of the risk measures and these are BL3_RT4 and BL4_RT4, where BL3 refers to “Retail Banking”, BL4 to “Commercial Banking” and RT4 to “Clients, Products and Business Practice”.

Both have a similar associated Capital at Risk, however, the difference between them lies in their charges for Expected Loss and Unexpected Loss. BL3_RT4 has a higher Expected Loss and BL4_RT4 a higher Unexpected Loss. The latter is due to the fact that the empirical Severity Distribution of BL4_RT4 has greater asymmetry to the right and greater kurtosis that can be appreciated in its heavier right tail, which leads to a greater Unexpected Loss (see *Figure 5*).

Another conclusion drawn from the fact that both Cells share the same Risk Type, RT4, is that this risk is the most damaging because it causes greater operational losses, leading to a higher Capital charge.

Given that companies have limited budgets to establish risk mitigation plans, in this Paper it is proposed that this budget should be allocated based on the Expected Loss, assigning a larger budget to the Cell with the highest

Expected Loss, in this case BL3_RT4. Point out that this budget will be additional to the Capital assigned to the Cell. The reason for this proposal is that the Expected Loss by definition is the potential loss with the highest probability of occurrence, therefore, it is essential to apply risk mitigation plans since these losses are very likely to occur.

6.2. Evaluation of the results of the total risk measures of the Entity

In this section the total risk measures are evaluated.

Risk Appetite had been defined as the amount of risk that an entity is willing to seek and assume in pursuit of its objectives, and Capital at Risk had been defined as the amount of money necessary to cover Expected and Unexpected Losses. Once these two definitions have been established, another two are introduced, the Risk Tolerance and Risk Capacity. The first refers to the acceptable level of risk variation and the second to the maximum amount of risk that the entity is capable of bearing.

Given these definitions, the Risk Capacity would coincide with the Capital at Risk and the Risk Tolerance would be equal to the difference between the Risk Capacity and the Risk Appetite.

Figure 16 sows a representation of the Risk Appetite, Risk Tolerance and Risk Capacity measures in the Aggregate Loss Distribution of the Entity SA.

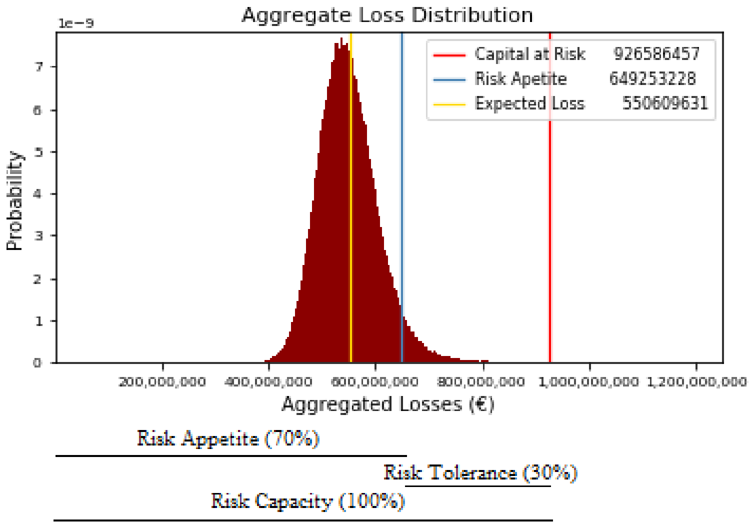

Figure 16. Risk Appetite, Risk Tolerance and Risk Capacity of the Entity SA

Analysing the graph, it can be appreciated that the magnitude of Risk Tolerance reaches 30% of the Risk Capacity. This percentage is considered quite high because it means that the Entity can bear a 30% risk in addition to the risk it was seeking to assume (Risk Appetite).

This allows to conclude that the proposed model has a notable degree of conservatism while improving the Standard Formula as it captures the entity's risk profile.

7. Conclusions

The proposed advanced model to quantify the Capital necessary for covering losses due to operational risk using a top-down approach has proven to be a robust model based on the Loss Distribution Approach (LDA), on the Monte Carlo simulation and on the Value at Risk measure.

The methodology captures the true risk profile of the Entity using only internal loss data, as these are already clearly linked to the Entity's business activities, technological processes and risk management procedures. It relegates the use of external loss data and scenarios, since the former, as they come from other financial entities, have a totally different nature and therefore do not reflect the risk profile of the Entity in question, and the latter are invented situations, which implies that both the Severity and Frequency data are unreliable, not making it clear to what extent they feed or contaminate the sample.

It solves the data scarcity problem by modelling jointly all the loss data, resulting in a large enough sample that allows capturing the true behavior of the data using parametric distributions for Severity and Frequency of few parameters without having to relapse into overfitting problems.

It achieves the maximum level of granularity a posteriori, since once the Distribution of Aggregate Losses of the Entity is obtained and with it the total risk measures are calculated (Capital at Risk, Risk Appetite, Expected Loss and Unexpected Loss), these are disaggregated into the Basel Cells.

Using the Monte Carlo simulation, the subjectivity factor is reduced to a minimum since a million possible losses are generated randomly through the convolution of the random variable of Severity with itself a number of times given by the random variable of Frequency. This convolution is necessary to combine discrete random variables from the Frequency Distribution and continuous random variables from the Severity Distribution.
It proves to be a model with a notable degree of conservatism because it is based on a high percentile for the Value at Risk measure, specifically, on the 99.90% percentile for Capital, in addition to it, the Risk Tolerance reaches 30% on the Risk Capacity with which the entity can bear a 30% additional risk to the risk it was seeking to assume (Risk Appetite), and what is more important, Capital covers expected and unexpected losses.

It allows identifying the Cells with the highest operational risk presented by the Entity, information that can be used for the provision of risk mitigation plans, assigning a greater budget to those Cells with the highest Expected Loss since, by its definition, it is the potential loss with the highest probability of occurrence.

Other conclusions obtained during the implementation were that extreme observations, both in Frequency and Severity, have a high impact on the Capital charge.

Therefore, it is concluded that the advanced model proposed proves to be a robust model capable of giving reliable estimates of Capital at Risk and Risk Appetite.

8. Bibliography

Basel Committee on Banking Supervision. (2011). *Principles for the Sound Management of Operational Risk*. Bank for International Settlements.

Carrillo, S., & Suárez, A. (2006). *Medición efectiva del riesgo operacional*. *Estabilidad financiera*, 11, 61-90.

Comité de Supervisión Bancaria de Basilea. (2006*). Observed range of practice in key elements of Advanced Measurement Approaches (AMA)*. Bank for International Settlements.

Day, N. (1969). *Estimating the components of a mixture of normal distributions.* Biometrika, 56(3), 463-74. DOI:10.1093/biomet/56.3.463

Fernández-Laviada, A. (2010). *La gestión del riesgo operacional. De la teoría a su aplicación*. ISBN: 9788495058652

Ferreras Salagre, A. (2007). *Riesgo operacional. Enfoque de distribución de pérdidas en la práctica*. Universidad Autónoma de Madrid. Madrid.

Frachot, A., Georges, P., & Roncalli, T. (2001). *Loss distribution approach for operational risk.* SSRN Electronic Journal, DOI:10.2139/ssrn.1032523

González, E. (2020). *Modelo avanzado de cálculo de Capital por riesgo operacional* (Trabajo Fin de Máster). Universidad Carlos III de Madrid.

Holgersson, M., & Jorner, U. (1978). *Decomposition of a mixture into normal components: a review.* International journal of bio-medical computing, *9*(5), 367–392. DOI:10.1016/0020-7101(78)90044-2

Tan, W. Y., & Chang, W. C. (1972). *Some comparisons of the method of moments and the method of maximum likelihood in estimating parameters of a mixture of two normal densities.* Journal of the American Statistical Association*,* 67(339), 702-708. DOI:10.1080/01621459.1972.10481282