Lisbon School of Economics & Management
Universidade de Lisboa

MASTER

MATHEMATICAL FINANCE

MASTER´S FINAL WORK

INTERNSHIP REPORT

AN EMPIRICAL APPROACH TO QUANTIFY MINIMUM CAPITAL REQUIREMENTS FOR OPERATIONAL RISK

MANUEL MARIA AVILLEZ ATAÍDE OLIVEIRA MONTEIRO

OCTOBER-2022

MASTER

MATHEMATICAL FINANCE

MASTER´S FINAL WORK

INTERNSHIP REPORT

AN EMPIRICAL APPROACH TO QUANTIFY MINIMUM CAPITAL REQUIREMENTS FOR OPERATIONAL RISK

MANUEL MARIA AVILLEZ ATAÍDE OLIVEIRA MONTEIRO

SUPERVISION:

DR. JOSÉ EDUARDO CRUZ
PROF. DOUTOR ONOFRE ALVES SIMÕES

OCTOBER - 2022

Acknowledgements

I want to thank my internship supervisor Dr. José Eduardo Cruz for all the help, support, advice, patience, and availability during my project in KPMG. I also want to thank:

My college supervisor Prof. Dr. Onofre Alves Simões for accepting my invitation, the advice towards my report and the sympathy in his words;

Dr. Pedro Loureiro for giving me the opportunity to start an internship in KPMG and all the understanding and flexibility towards my working hours.

All the people from the Risk and Management Consulting department for all the guidance and help they gave me to start this opportunity, for their companionship and for including me in other events;

My family for all the love and support, for putting up with me during these months of hard work and for their kindness;

To my friends for supporting me, for hanging out with me out of working hours and for understanding my effort for this internship that could keep me busy sometimes without judging.

Resumo

Este trabalho é o relatório de um estágio de seis meses levado a cabo na consultora KPMG Advisory, no *Department of Risk Consulting*.

Historicamente, as instituições financeiras têm reconhecido a importância que a quantificação dos chamados requisitos mínimos de capital tem para a gestão de risco. Com efeito, estes requisitos estabelecem limites exigidos por reguladores corporativos, corretores ou agências, na realização de operações financeiras, com a finalidade de proteger tais instituições da falência, insolvência ou outras situações de crise. A União Europeia lançou vários documentos onde dá orientações para a minimização do risco de perda de capital, destacando-se o Regulamento de Requisitos de Capital (CRR), cujo objetivo é guiar os bancos na implementação de um conjunto de ações normalizadas, para gerir riqueza e simultaneamente evitar a ocorrência de crises agravadas.

Durante o meu estágio participei num projeto de Modelo de Risco Operacional destinado a apoiar as decisões do Processo de Autoavaliação da Adequação do Capital Interno (ICAAP) de um banco português. No seu decurso fui chamado a programar um algoritmo para efetuar uma simulação de Monte-Carlo para as perdas esperadas no âmbito daquele risco, que seriam posteriormente traduzidas em termos de requisitos mínimos de capital. O modelo baseia-se essencialmente no CRR, mas incorpora também algumas indicações da Abordagem de Medidas Avançadas (AMA).

Para obter os resultados, foi necessário efetuar dois tipos de agregações das unidades de medida de risco (ou células de risco), sob indicação do banco. As agregações escolhidas corresponderam a resultados diferentes, sendo as causas das diferenças analisadas e discutidas em função da frequência e gravidade dos eventos.

Palavras-Chave: Monte Carlo, Risco Operacional, Requisitos Mínimos de Capital, Distribuições Corpo-Cauda

Abstract

This work is the report of a six-month internship carried out at KPMG Advisory, in the Department of Risk Consulting.

Historically, financial institutions have recognized the importance that calculating the so-called minimum capital requirements has for risk management. Indeed, these requirements establish limits required by regulators, brokers or agencies, in carrying out financial operations, in order to protect such institutions from bankruptcy, insolvency or other crisis situations. The European Union has released several documents which provide guidelines for minimizing the risk of capital loss, in particular the Capital Requirements Regulation (CRR), whose objective is to guide banks in the implementation of a set of standardized actions to manage wealth and simultaneously to prevent the occurrence of aggravated crises.

During my internship I participated in an Operational Risk Model project, aimed at supporting the decisions of the Internal Capital Adequacy Assessment Process (ICAAP) of a Portuguese bank. I was asked to program an algorithm to perform a Monte Carlo simulation for the expected losses associated to the Operational Risk, which would later be translated into terms of minimum capital requirements. The model is essentially based on the CRR, but also incorporates some indications of the Advanced Measures Approach (AMA).

To obtain the results, it was necessary to carry out two types of aggregation of the risk measurement units (or Risk Cells), as indicated by the bank. The aggregations chosen would correspond to different results, and the causes of the differences were analyzed and discussed based on the frequency and severity of the events.

Key Words: Monte Carlo, Operational Risk, Minimum Capital Requirements, Body-Tail distributions

Index

1. Introduction ... 1

1.1. Description of the Organization ... 2

1.2. Internship Plan ... 3

2. Theory ... 5

2.1. Operational Risk ... 5

2.2. ICAAP ... 7

3. Methodology ... 8

3.1. Operational Risk Model ... 8

3.2. Risk Cells ... 9

3.3. Probabilistic Models for Frequency and Severity of Losses ... 10

3.3.1. Frequency Model ... 10

3.3.2. Severity Model ... 11

3.4. Monte Carlo Simulation ... 14

4. Application ... 19

4.1. Database and Software ... 19

4.2. Aggregation 1 ... 20

4.3. Aggregation 2 ... 24

4.4. Results and Discussion ... 27

5. Conclusion ... 30

References ... 31

Appendix ... 34

List of Tables

Table 4.1. Frequency Parameter for Aggregation 1……………..……..........…...………………21
Table 4.2. Finding the Body-Tail threshold for Risk Cell ELgr, Aggregation 1.........................22
Table 4.3. Finding the Body-Tail threshold for Risk Cell EL7, Aggregation 1..........................22
Table 4.4. Severity Parameters for Aggregation 1………………..……..........…...…...................23
Table 4.5. Monte Carlo Results After 1000 Iterations, Aggregation 1......................................23
Table 4.6. Monte Carlo Results After 10000 Iterations, Aggregation 1....................................23
Table 4.7. Monte Carlo Results After 100000 Iterations, Aggregation 1..................................24
Table 4.8. Monte Carlo Results After 1000000 Iterations, Aggregation 1................................24
Table 4.9. Frequency Parameters for Aggregation 2…………………..………..………......……26
Table 4.10. Severity Parameters for Aggregation 2………………………………….................26
Table 4.11. Monte Carlo Results After 1000 Iterations for Aggregation 2.................................26
Table 4.12. Monte Carlo Results After 10000 Iterations for Aggregation 2..............................27
Table 4.13. Monte Carlo Results After 100000 Iterations for Aggregation 2............................27
Table 4.14. Monte Carlo Results After 1000000 Iterations for Aggregation 2..........................27

List of Figures

Figure 3.1. Aggregation of Event Types 9
Figure 3.2. Frequency and Severity Distributions..10
Figure 3.3. Scheme of the Body-Tail Distribution..11
Figure 3.4. Empirical Distribution of Losses..13
Figure 3.5. Simulation of the Risk Cell's Frequencies..15
Figure 3.6. Algorithm to Simulate Body Loss Distribution..16
Figure 3.7. Lognormal Distribution Density Function..17
Figure 3.8. Algorithm to Simulate Tail Loss Distribution..17
Figure 4.1. Distribution of Events Among the Seven Categories – Aggregation 1....................20
Figure 4.2. Frequency of Events per Quarter for Aggregation 1..21
Figure 4.3. Distribution of Events Among the Seven Categories – Aggregation 2....................25
Figure 4.4. Frequency of Events per Quarter for Aggregation 2..25

1. Introduction

Within the scope of the curricular program of the Master's in Mathematical Finance, at Lisbon School of Economics and Management, Universidade de Lisboa, this is the report of a 6-month internship at KPMG Advisory, in the Department of Risk Consulting.

Despite of being one of the Big Four consulting firms, with a huge spread over the world and a vast range of opportunities for recent graduates, KPMG is aware of the main barriers that they face when entering the job market. To better integrate them, the more experienced staff is prepared to help newcomers with any type of problems, inserting them into projects that are suitable to their profiles, as well as being available to clarify any doubts along the internship.

During this work experience, I was involved in a project and helped some colleagues of my department team in Financial Service – Risk Consulting, in tasks where I could be useful.

The main project of my internship was the development of an Operational Risk (OR) model (Jöhnemark, 2012), created with the purpose of quantifying the expected losses of a bank associated to OR, through Monte-Carlo simulation, and distributing them by tail or body. The model uses statistical distributions, like Poisson, Negative Binomial, Binomial and Lognormal, to obtain a more suitable capital requirement considering the bank's portfolio. The ultimate objective of this model is to measure the Minimum Capital Requirements (MCRs) (Anghelache, Olteanu, & Radu, 2010) that may influence the ICAAP decisions about operations made by the bank (De Jonghe, Dewachter, & Ongena, 2020). It is an important framework because financial institutions are not totally certain about how much capital is necessary to perform all their financial operations, whether loaning money for an entity, issuing derivatives, investing in companies, paying salaries, or improving the building facilities. Therefore, through Operational Risk management, banks have information about their limits of wealth capacity to manage more efficiently the costs of financial operations, without putting the other counterparts in risky situations.

During the first three semesters of my master's program, I studied a variety of subjects that prepared me to work in this project, like Programming Techniques (that helped me to develop the algorithm codes for Monte Carlo simulation) and Financial Econometrics (which gave me the tools to interpret the parameters of the fitted distributions and to test their adequacy). Other subjects from my master's curricular plan, such as Interest Rate and Credit Risk Models, for example, will be quite useful for Risk Consulting, due to their content related with probability of default, Value-At-Risk models, or the Merton model, which are concept and tools that the Department of Risk Consulting uses very often.

1.1. Description of the Organization

KPMG is one of the worldwide Big Four multinational consulting firms, offering several professional services in three areas: Audit and Assurance, Tax and Advisory.

The Audit and Assurance services guarantee the reliability of the information obtained in capital markets and provided by investors; the Tax services provide legal and regulatory assistance, which deals with the policies, laws and bureaucracy related with the company; the Advisory services provide new solutions, tools and take the initiative of several projects in cooperation with other firms.

There are KPMG headquarters spread over 147 countries and currently there are around 227 000 working professionals of all the three areas. My working office is in Lisbon, building FPM 41, Avenida Fontes Pereira de Melo, number 41, and I am working in the Risk Consulting - Financial Services team department, which belongs to the Advisory services, and it is on the same floor as the Management Consulting and Deal Advisory teams, both from the same area.

The department of Risk Consulting is responsible to deal with large quantities of information, to manage several types of risk and solidify confidence in future decisions of various business executives, when these are upon the moment at performing an action. This sector deals with risk management of several types, like tax fraud, cyberattacks, regulatory compliance, data violation, structure and credit risk models, capital efficiency and corporate governance.

Also, this department has a very high demand for applied mathematicians, due to their knowledge of statistics and probability as well as a basic understanding of programming and working with databases, which are not yet very usual to find in consultancy members with other academic backgrounds. Another positive factor is their methodological thinking, which helps to reach stronger and well-based solutions and manage different variables with time, based on statistical tools adapted to time variation. Therefore, they are recognized as very well-prepared employees, who not only have different skills that contribute to the project, but are also prepared to learn more and constantly develop their performance with an abstract and refined logical reasoning, to absorb new information, and independence, to offer new solutions by themselves.

My KPMG's supervisor is named José Cruz, a Manager in the Risk Consulting department. He was responsible to introduce me to a project he was inserted at the beginning of the internship along with other employees of the same team department, and so he helped me to adapt to the

fasten rhythm of KPMG workers. He was available to clarify all my doubts related to the project and supervised me, to make sure I was doing well the activities he requested.
After the pandemic, the company started to be more flexible, by establishing a hybrid schedule, meaning that workers had specific days of the week to work at home, while the rest of the week was to work at the office.

1.2. Internship Plan

Before initiating the internship, KPMG organized an activity timeline, covering the six months of the internship. It was designed to be a guideline, planning the future tasks to accomplish. According to the internship plan in Appendix 1, my main activities were:

- Identification of the main applicable regulatory requirements for the project;
- Analysis of the key points that were addressed into risk quantification;
- Analysis of the main difficulties about measuring risk;
- Development of metrics to measure the detected risks.

The cells in grey highlight the activities made in each month. Since the project of the Operational Risk Model lasted around two months, to fulfill the remaining time I had to perform accessory tasks of other projects, demanded by other superiors from the same department team. So, the internship plan activities did not follow the same order as it is presented in Appendix 1, even though they were already included in the projects I have participated. I have also helped other colleagues of the department while preparing the internship report. In fact, in addition to the OR project, I supported and helped temporarily in three other projects.
The first one was from KPMG in Germany, and it was about Physical and Transition Climate Risks Prototypes, to assess the Expected Credit Loss (ECL) until the year 2100, related with possible flood disasters and their damage impact, as well as the climate scenarios from the Network for Greening the Financial System[1] - for further details, see (Luo, et al., 2021) and KPMG's paper[2].
In that project, my tasks were to interpret two Python program outputs associated to the Physical and Transition Risks, explaining the equations behind the calculations performed by the programming code, and analyzing the plots, how and where the data was coming from, and finally relating all the results to reach the required ECL.
The second one was a study about retention of term deposits, for liquidity risk management purposes. The client is a financial institution in Mozambique, and I helped a colleague to create

[1] https://www.ngfs.net/en
[2] What's the impact on expected credit losses? - KPMG Global (home.kpmg)

an optimal segmentation analysis on term deposits. In order to achieve that, I had to perform a historical analysis, to see which products have a heavier representation on the term deposits, also had to do a parametrization of the renovations and early withdrawals and a correlation analysis, and finally had to calculate the limits of a 95% confidence interval to some groups of products that have similar names, to check the possibility of merging them.

The third one was a project in the scope of Internal Audit under the IFRS 9 norm for financial instruments, cf. PWC's document (Audit Services | IFRS 9, 2017). I joined a department team of Internal Audit and my tasks were programming in SAS language with one member of the team for a database from the customer. I mapped stages 1, 2 and 3 to the various customers' contracts that had notifications related with credit risk events, finding triggers that warn the bank about possible default cases, then validating them to make sure the database had the right trigger values, and finally investigating how the SAS database was calculating the ECL for stage 3, according to the Bank for International Settlements' guidelines (FSI | IFRS 9 and expected loss provisioning - Executive Summary).[3]

Therefore, it is reasonable to say my internship was both academic and professional and allowed me to be familiar with a number of different tools and ways of proceeding and reasoning the tasks. Additionally, it presented various opportunities to participate in projects of the company, which clarified me the issues and applications of Risk Analysis.

As it was stated at the beginning of this chapter, the internship timetable worked as a guideline for the topics I would have to surpass for the various projects, but that did not necessarily mean the obligation to follow each activity in a certain month.

The progression of the paper is as follows. In Chapter 2, the concepts of Operational Risk and ICAAP will be presented, Chapter 3 covers the model and the methodology, in Chapter 4 the methodology is applied to two Risk Cell aggregation types, Chapter 5 presents Results and Discussion about the outcomes obtained in Chapter 4, and Chapter 6 contains the main conclusions and some suggestions to improve this framework.

[3] https://www.bis.org/fsi/fsisummaries/ifrs9.pdf

2. Theory

2.1. Operational Risk

The definition of Operational Risk is «a risk of loss born from failures and misalignment of internal processes, people, or systems, as well as exposition to external events that include legal risks that cause or might have caused material losses or decrease for the shareholders (The European Parliament, 2013, p. L 176/22).»

Any loss resulted from an OR event related with market risk, like trade losses or gains amplified or reduced by adverse fluctuations, or through market value with origin in operational errors, is recognized as an OR event and therefore must be included into the OR model, as well as any loss resulting from legal action, whose genesis has also origin in OR.

According to a KPMG internal document[4], the Operational Risk management framework is divided into three main parts:

I. **Strategy and Risk Appetite;**
II. **Organization and Governance;**
III. **Management Instruments.**

I. Strategy and Risk Appetite

Unlike financial risks, such as credit, legal or market types, the OR has a different nature, with some particularities regarding the traditional view of the risk appetite.

While in the financial risks there is a balance between profitability and risk (the risk/reward ratios)[5], in the operational type risk no expected reward is assumed, because it depends on the firm's activity. This contradicts the purpose of the financial risk appetite, which is the risk of loss that the bank assumes in exchange for an expected reward. Instead, the purpose of the Operational Risk appetite is to find the optimal point where the marginal expense equals the marginal reduction of expected losses.

II. Organization and Governance

In risk management, there are three lines of defense that characterize its process:

— Management control;
— Supervision functions established by management on compliance and risk control;

[4] Internal document from KPMG (2021) | *Pilar I: Solvência, TRIM e Risco Operacional*
[5] Risk/Reward Ratio: What It Is, How Stock Investors Use It (investopedia.com)

— Independent assurance provided by internal audit.

Risk managers usually belong to the first line of defense, and they are responsible for managing the tasks of OR, like identification, mitigation, or evaluation of loss events. Also, there are risk coordinators that support risk managers on their daily duties and help them to make the interlocution with the second line of defense.

For several types of ORs, there are the Specialized Control Functions that are responsible to provide a global vision of the exposition to the most relevant typologies of OR and mitigate it through the application of controls. These functions are easy to find in support areas like compliance, IT or risk analysis, where all of them belong to the second line of defense.

The third line of defense basically verifies the effectiveness of the previous two lines to achieve the objectives set by the risk managers, since it corresponds to an independent process from the Operational Risk management and warrants the efficiency of the framework by mitigating all the risks.

III. Management Instruments

To perform an efficient management of the OR, it is necessary to have specific instruments and transversal elements, like culture and communications, norms and policies, internal control or backtesting, for a clean execution of the model and to ease risk control and management.

They are used in three different phases of the structure for OR management:

- First, the Identification and Evaluation that deals with Risk Control Self-Assessment[6], the internal database of events, scenario analysis and makes the evaluation of new products along with transition management;
- Second, the Monitorization through Operational Risk Indicators, see (Davis & Haubenstock, 2002), to check and validate occurrences like loss events;
- Third, the Mitigation of OR, which improves the business continuity plan and the risk transfer.

According to the guidelines of line b) number 3, article 322 from the CRR, hedging the OR losses, although always registered in the internal database, must not be subject to own fund requirements, since it is supposed to avoid duplications of capital requirements.

[6] The Methodology Behind Risk and Control Self-Assessment (theglobaltreasurer.com)

2.2. ICAAP

According to Basel II, the second pillar relies on supervisory action (Rochet, 2004), which is responsible for the regulatory response for the capital requirements and the development of the ICAAP report.

ICAAP is the set of internal procedures and systems that guarantee the optimal allocation of capital resources for the bank in a long-term horizon to cover all its material risk impacts. It aims determining the economic capital, which is the capital required to cover all risks that are estimated, using the bank's internal risk models. Its main purpose is to ensure a suitable relation between the bank's overall capital and its level of risk exposition (Farid, 2010).

The documentation of an ICAAP report should:

- Inform the bank's board of directors about the continuous evaluation of all the firm's risks;
- Inform the board of directors and the senior management about the main results of the risk assessments, how the firm wants to mitigate those risks and how much future capital is needed, as well as explaining its consequences;
- Explain the Internal Capital Adequacy Assessment Process made to the supervisor of the bank.

3. Methodology

3.1. Operational Risk Model

The OR model developed in the context of the project is based on articles 312-324 of the European Capital Requirements Regulation 575/2013, considering some requirements of the Process and Internal Data components from the Quantitative Standards (article 322 (2-3)), see Appendix 2.

The institutions can use some or all the parts of the Advanced Measure Approach (articles 321 and 322), if they "notify the competent authorities of all changes to their Advanced Measurement Approaches models" (article 312, paragraph 3). It includes an internal database with historical information of operational loss events, structured in a way that enables to select the events according to the risk category. The whole framework is based on a solid data collection procedure, which aims to ensure the quality and completeness of the internal database, since it is used to estimate components for the expected losses of the distribution.

The risk categories used in the model are:

1. Internal Fraud (EL1);
2. External Fraud (EL2);
3. Employment Practices and Workplace Safety (EL3);
4. Clients, Products & Business Practices (EL4);
5. Damage to Physical Assets (EL5);
6. Business Disruption and System Failures (EL6);
7. Execution, Delivery & Process Management (EL7).

Despite the existence of loss events from all these risk categories, due to lack of data in some of them, an aggregation was performed to strengthen the results. For this model, only events occurred in Portugal were considered, according to the bank's structure and their contribution to the Relevant Indicator of the Standardized Approach, following the Basel Committee document on (OPE Calculation of RWA for Operational Risk, 1999, pp. 11-19) and (The European Parliament, 2013, pp. 196-197).

To obtain the components of the model we retained all the events related to Operational Risk with loss amounts after direct recovers of at least 250 euros, such that:

$$\text{Liquid Loss Amount} = \text{Gross Loss Amount} - \text{Direct Recovers Amount} \geq 250 \quad (1)$$

Regarding recoveries, only direct recoveries were considered in the calculation of the Liquid Loss Amount, because they result from natural causes that consequentially reflect in the amount, which enabled the bank to recover partially or totally the Gross Loss Amount.
The indirect recoveries are related with the time deferral, which sometimes extends from the payment of the compensation by the insurer, the lack of evidence from general conditions into current policies and, at this stage, the preference for a more conservative approach to the model[7].
According to the European Bank Authority document EBA/RTS/2015/02, the construction of the frequency and severity distribution datasets must be based on the accounting date or the detection date of the event, to stabilize the historical database for past periods.

3.2. Risk Cells

The Risk Cells described in Section 3.1 (EL1-EL7) were defined in the model at the level of the risk categories specified in article 324 of the CCR (see Appendix 3). Due to the lack of frequency and severity of collected events in some of these risk categories, it was decided to merge them. There are categories that have enough events and so can be considered as a unique Risk Cell, while others with few events must be aggregated.
This does not mean the categories with sufficient events must be automatically considered as individual Risk Cells, since the aggregation attends to the preservation of homogeneity of the data in each resulting Risk Cell.
The scheme of aggregation choices follows the algorithm below:

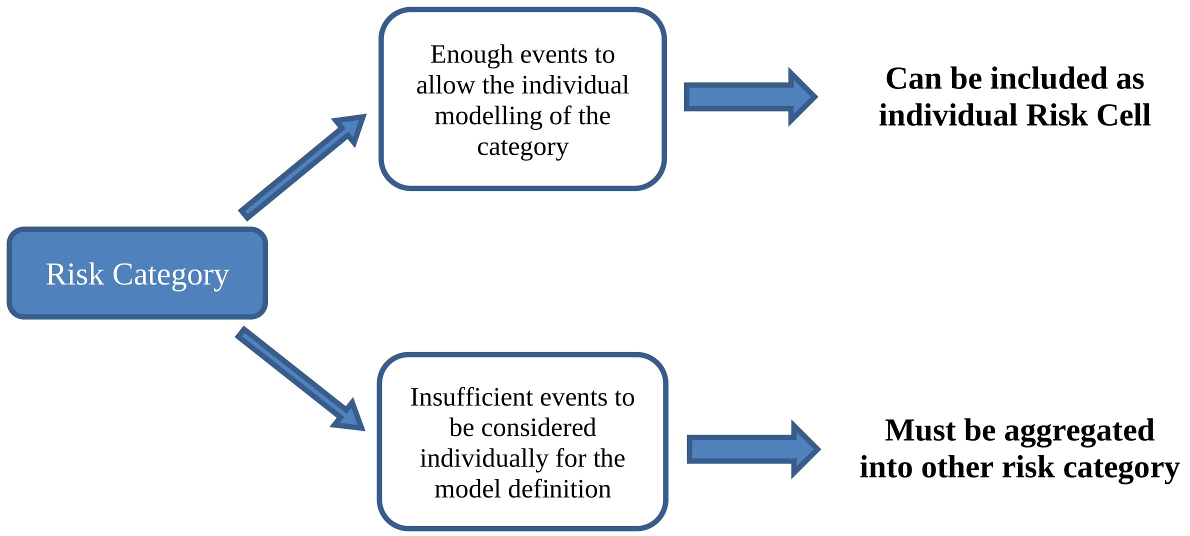

Figure 3.1. Aggregation of Event Types

[7] What is indirect and consequential loss? - Harper James

3.3. Probabilistic Models for Frequency and Severity of Losses

In this section, we present a description of the procedures to approach the two main components of the OR model: the frequency and the severity of the losses.

The frequency is the number of occurrences in a specific time period, which can be at least five years[8], and the severity represents the material financial impact (in this case the total monetary amount of losses), as the following graphs illustrate.

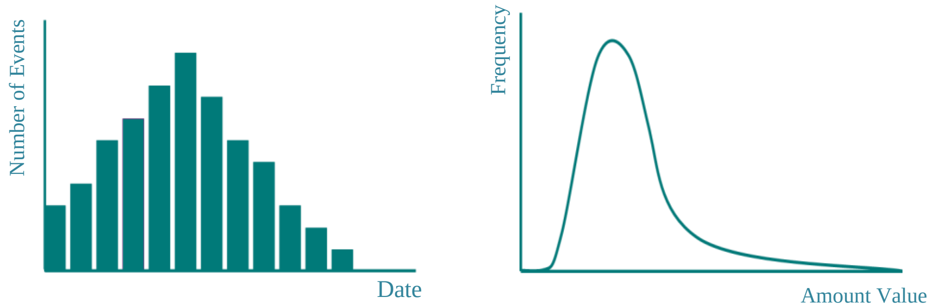

Figure 3.2. Frequency and Severity Distributions

For the frequency variable (left graph), two discrete probability distributions, the Poisson and the Negative Binomial, were selected (Jöhnemark, 2012). For the severity variable (right graph), we used one continuous probability distribution, the Lognormal (Bermúdez, 2015).

These distributions will enable to determine the input parameters for Monte Carlo simulation and therefore to reach the results for the Minimum Capital Requirements.

3.3.1. Frequency Model

Based on a dataset with internal losses, the parameters of the distributions have to be estimated, since these estimates are required to set the model.

First, for each Risk Cell, the number of loss events in each quarter of the year is counted, and then the Mean and the Variance are calculated.

Second, the Critical Ratio (Variance/Mean) 'rule' is applied. This indicator is quite important to decide which is the 'right' distribution[9] to be used later for the Monte Carlo simulation of the respective risk measure. According to the 'rule':

- If Critical Ratio $> 2 \rightarrow$ Negative Binomial Distribution;
- If Critical Ratio $\leq 2 \rightarrow$ Poisson Distribution.

[8] Article 322 paragraph 3a) from Regulation 575/2013 (CRR)

[9] The choice of the distribution to model the frequencies loses relevance when considering the percentile to be used for the loss distribution.

This ratio is used since Poisson distribution has the mean equal to the variance, therefore for its usage, the observed values also must be similar, and the ratio should return a value around 1.
So, to model the frequency for each Risk Cell, three parameters are potentially relevant:

- 'Lambda' (λ), the Poisson parameter which is, in this case, equal to the mean;
- 'Size' (r), the first Negative Binomial input estimated through maximum likelihood and Brent's algorithm (Brent, 1971);
- 'Mu' (p), the second Negative Binomial input estimated through maximum likelihood and Brent's algorithm.

3.3.2. Severity Model

Severity is composed of two parts: Body and Tail, separated by an amount, the Threshold. The Body is defined as the major part of the severity distribution, where all the losses have an amount less than or equal to the Threshold; the Tail corresponds to the upper part of the distribution, representing all the loss amounts higher than the Threshold, as represented in Figure 3.3.
Again, based on a dataset with internal losses, the parameters of the severity distribution have to be estimated, since these estimates are required to set the model.

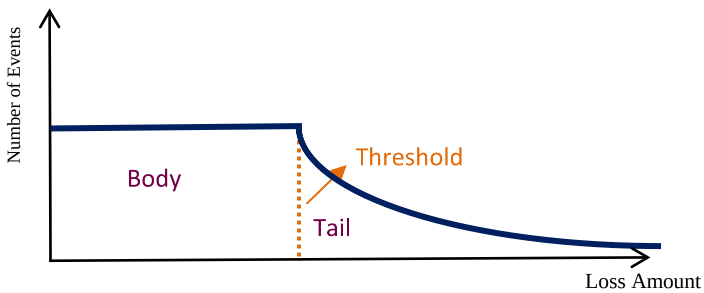

Figure 3.3. Scheme of the Body-Tail Distribution

According to Basel II, "a bank must be able to demonstrate that its approach captures potentially severe 'tail' loss events" (see Appendix 4). So, it is possible to simulate the tail loss values through the Lognormal distribution as an approach to catch the events with the highest severity.
Therefore, for each Risk Cell, four quantities are relevant:

- The Threshold;
- The Mu, the first parameter of the Lognormal distribution;
- The Sigma, the second parameter estimated for the Lognormal distribution;
- The Bodyweight, the probability of the loss being assigned to the Body.

It is crucial to know the Threshold value, to calculate the remaining quantities, but that depends on the bank's inside information, resulting in a known or unknown number. If the bank does not request to model the severity with a specific threshold, it must be determined by some type of inspection. To do this, a table is constructed with fits and statistical experimental tests, for the available loss data. This table gathers all the information about the different experiments, which includes not only the Threshold value but also the Bodyweight, the Mu and the Sigma, sorted by Risk Cell, percentile of the losses and fitted distribution. So, from the practical perspective, the choice of the Threshold is based on selecting one line from that table for each Risk Cell, which relates:

- -The number of events considered for the Tail, which should be the minimum possible according to the bank's criteria.
- -The Mu and Sigma parameters are the highest possible, which sometimes don't increase proportionally with the percentile.

The lognormal distribution parameters are determined through maximum likelihood estimation for the events in the Tail, as follows:

- $$\hat{\mu} = \frac{\sum_i^N ln(X_i)}{N} \Longrightarrow \boxed{\text{Mu (parameter 1)}} \qquad (2)$$

- $$\hat{\sigma} = \sqrt{\frac{\sum_i^N (\ln(X_i) - \hat{\mu})^2}{N}} \Longrightarrow \boxed{\text{Sigma (parameter 2)}} \qquad (3)$$

$X_i = Value\ of\ the\ i^{th}\ loss\ event$

$N = Number\ of\ events\ in\ the\ Tail$

Finally, the Bodyweight is defined as the weight distribution before reaching the Threshold's quantile. It is obtained through the Empirical Cumulative Probability Function (ECPF) to find the weight of the loss distribution until the threshold value, see Figure 3.4

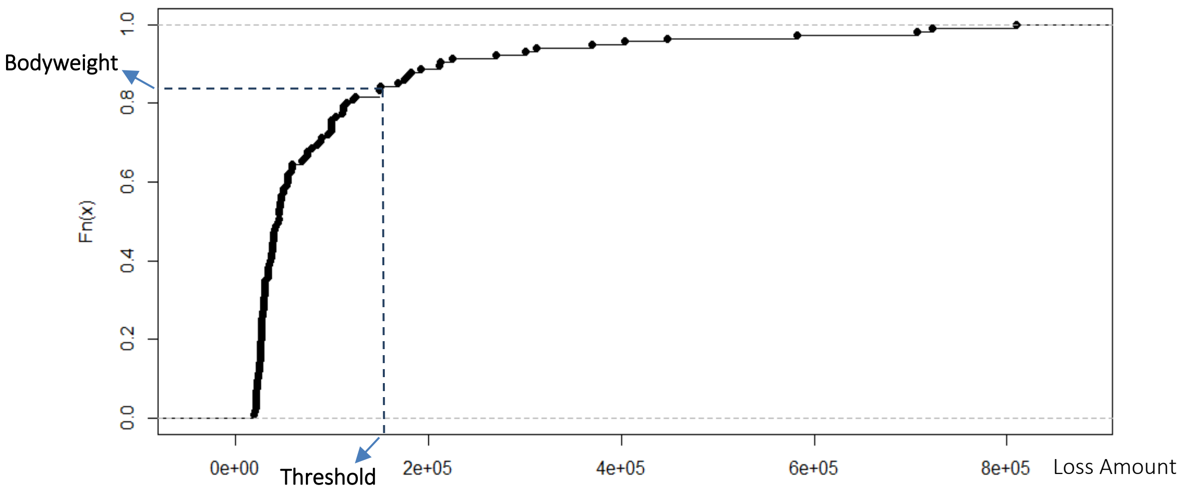

Figure 3.4. Empirical Distribution of Losses

To achieve the Bodyweights for all Risk Cells, outliers of the distribution are disregarded, that is, the loss values lower than the 1st percentile (P_1) or higher than the 99th percentile (P_{99}). This prevents unstable results or shocks in the calculations, and they will not be accounted for in the bodyweight estimation.

As it is a cumulative distribution, the loss values are sorted and summed until the closest value lower or equal to the correspondent Risk Cell's Threshold. Finally, that sum is divided by the total sum of event losses, and so it is obtained the bodyweight.

For each Risk Cell the bodyweight is given by:

$$\text{Bodyweight} = \frac{\sum_{i=1}^{n} l_i}{\sum_{i=1}^{N} l_i}, P_1 \leq l_i \leq \min\{\text{ Threshold}, P_{99}\}, \tag{4}$$

- $n = Number\ of\ Events\ between\ P_1\ and\ the\ Threshold$
- $N = Number\ of\ Events\ between\ P_1\ and\ P_{99}$
- $l_1 \leq l_2 \leq \cdots \leq l_N - Event\ Loss\ Amounts$

The bodyweight value increases when the Threshold gets higher, thus allowing more events to be considered from the Body in detriment of the Tail, as shown in Figure 3.4.

When the bank supplies the value of the Threshold, the procedure is the same as described above.

3.4. Monte Carlo Simulation

As it is usual with problems with this level of complexity and amount of data, the decision was to use the Monte-Carlo simulation technique to estimate the loss distribution for one year (see for instance Korn, Korn and Kroisandt (2010) for more details). The technique is expected to output a well-defined loss distribution with fairly accurate results. On the other hand, one well-known limiting aspect of this type of approach is that it is based on past information, and one must be careful when relying on previous results to forecast future losses.

In the following lines we show how the Monte Carlo method was adjusted to our OR model, including schemes, figures, and algorithms, to clarify its interpretation.

So, **for each Risk Cell**:

Step 1: Insert the input data:

- Parameters of the discrete distributions (Poisson and Negative Binomial) to simulate frequency, previously estimated as explained above;
- Critical Ratio, to choose the right distribution;
- Parameters of the Lognormal distribution to simulate severity (excess of losses beyond the Body-Tail threshold), previously estimated as explained above;
- Loss records;
- Required number of simulations.

Step 2: Generate a vector with dimension equal to the number of iterations (usually is 1000000 entries), where each entry corresponds to a *random variable of the uniform distribution between 0 and 1*, ($y_k \sim U(0,1)$), where k is the number of the iteration.

Step 3: Use the vector in Step 2 and the Generalized Inverse Transform Method (Korn, Korn, & Kroisandt, 2010) to simulate observations either from the Poisson distribution or the Negative Binomial Distribution, as appropriate (See Figure 3.5), where each observation is denoted X_k:

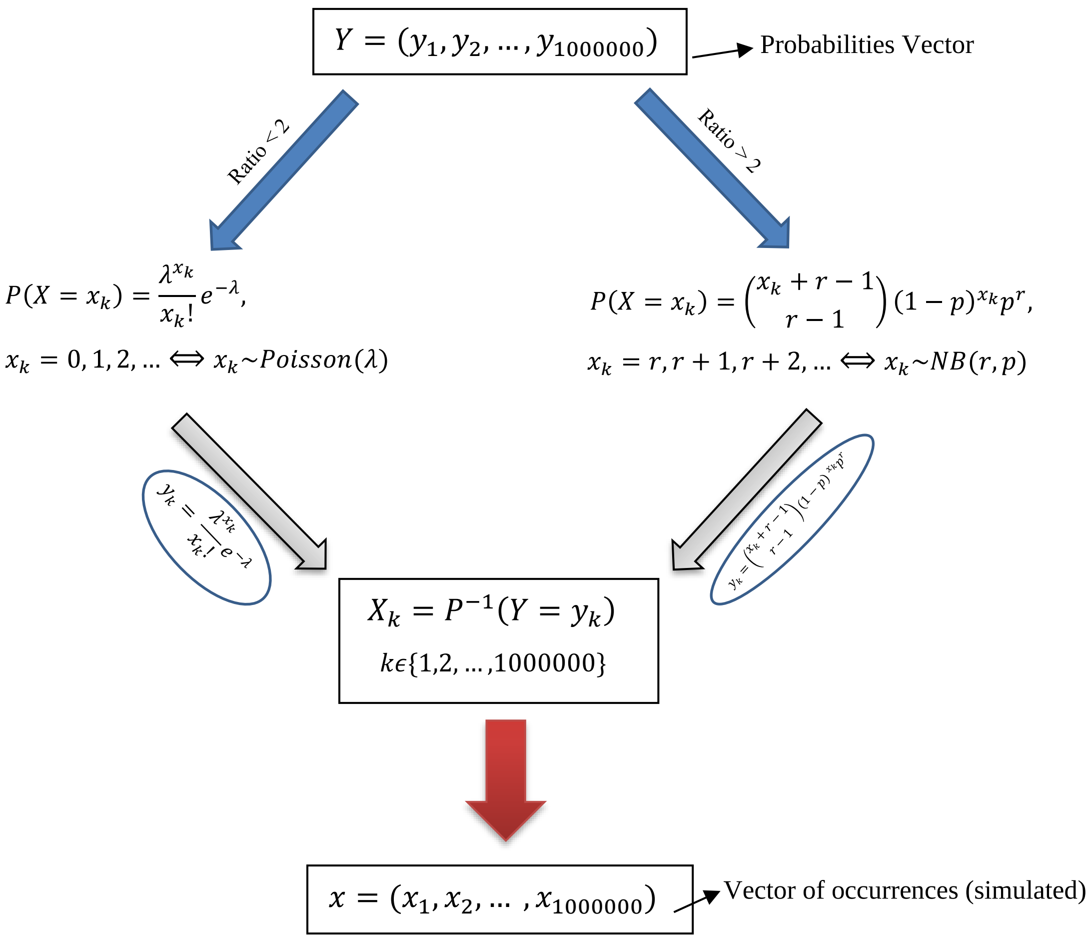

Figure 3.5. Simulation of the Risk Cell's Frequencies

Step 4: Decompose $x = (x_1, x_2, \dots, x_{1000000})$, the vector with the number of occurrences, as

$$(x_1, x_2, \dots, x_{1000000}) = (b_1, b_2, \dots, b_{1000000}) + (t_1, t_2, \dots, t_{1000000}), \quad (5)$$

where:

- $B = (b_1, b_2, \dots, b_{1000000})$ is the vector with the number of events that have loss amounts below the threshold (in the body of the distribution of losses);
- $T = (t_1, t_2, \dots, t_{1000000})$ is the vector with the number of events that have loss amounts at least equal the threshold (in the tail of the distribution of losses).

$(\forall b_i \in B,\ b \sim B(x, p))$, meaning that B is obtained multiplying x by the estimated percentage of claims in the body of the distribution (rounding to the nearest integer), and $\boldsymbol{T = x - B}$.

Knowing that each event is associated with a loss, new vectors must be defined:

- $L_k = (l_{1k}, l_{2k}, \dots, l_{x_k k})$ is the vector with the amounts of the x_k occurrences simulated in iteration k, $k = 1,2, \dots, 1000000$.
- $V =$ Value of the threshold.

Because of the decomposition operated to x it is necessary to split L_k in two vectors, the Body Losses vector ($L_k^{(B)}$ $with\ b_k\ entries$) and the Tail Losses vector ($L_k^{(T)}$ $with\ t_k\ entries$).

Step 5: Body loss distribution.

For each iteration $k = 1,2,\ldots,1000000$, the algorithm randomly selects b_k losses from the Body data (whose elements are the m observed losses with amounts below the threshold V) and adds them up, simulating this way the sum of the claims below the threshold (Body) for each iteration, see Figure 3.6 for an illustration.

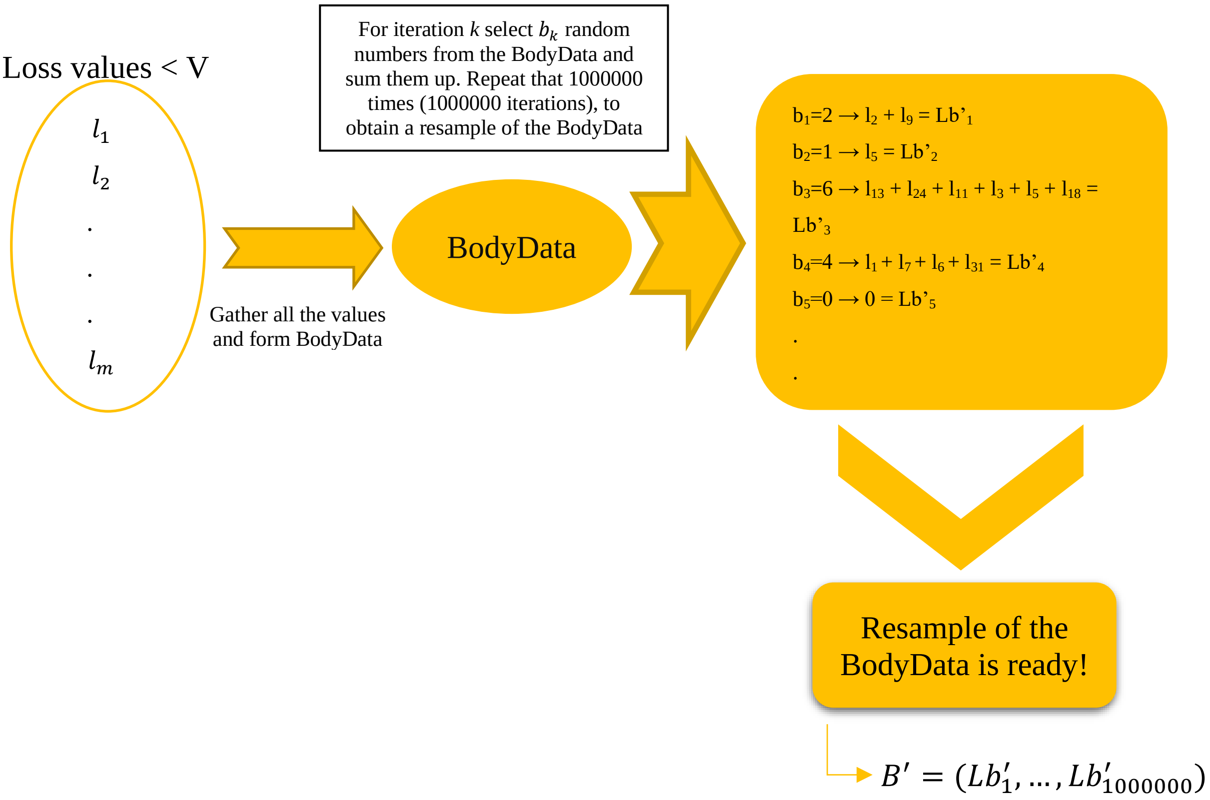

Figure 3.6. Algorithm to Simulate Body Loss Distribution

Step 6: Tail loss distribution

For each iteration $k = 1,2,\ldots,1000000$, the algorithm simulates t_k losses from the estimated Lognormal distribution and adds the Body-Tail threshold V, to each one of them. Next, it adds the t_k results up, simulating this way the sum of the claims over the threshold (Tail) for iteration k, see Figures 3.6 and 3.7 for an illustration. In other terms:

$$k = 1,2,\ldots,1000000.$$
$$i = 1,\ldots,t_k \qquad Lt'_{ki} = V + w_{k,i}, \text{ where } w_{k,i} \sim LogN(\underbrace{\mu}_{\text{Parameter 1}}, \overbrace{\sigma}^{\text{Parameter 2}}) \qquad (6)$$

Recall that the Lognormal distribution is defined by the following density function:

$$f(w,\mu,\sigma) = \frac{1}{w\sigma\sqrt{2\pi}} exp\left(-\frac{(\ln(w)-\mu)^2}{2\sigma^2}\right) \quad (7)$$

Figure 3.7. Lognormal Distribution Density Function

where μ and σ represent parameters 1 (Mu) and 2 (Sigma) of the distribution, respectively. The results for the Tail data are performed as shown in the scheme below.

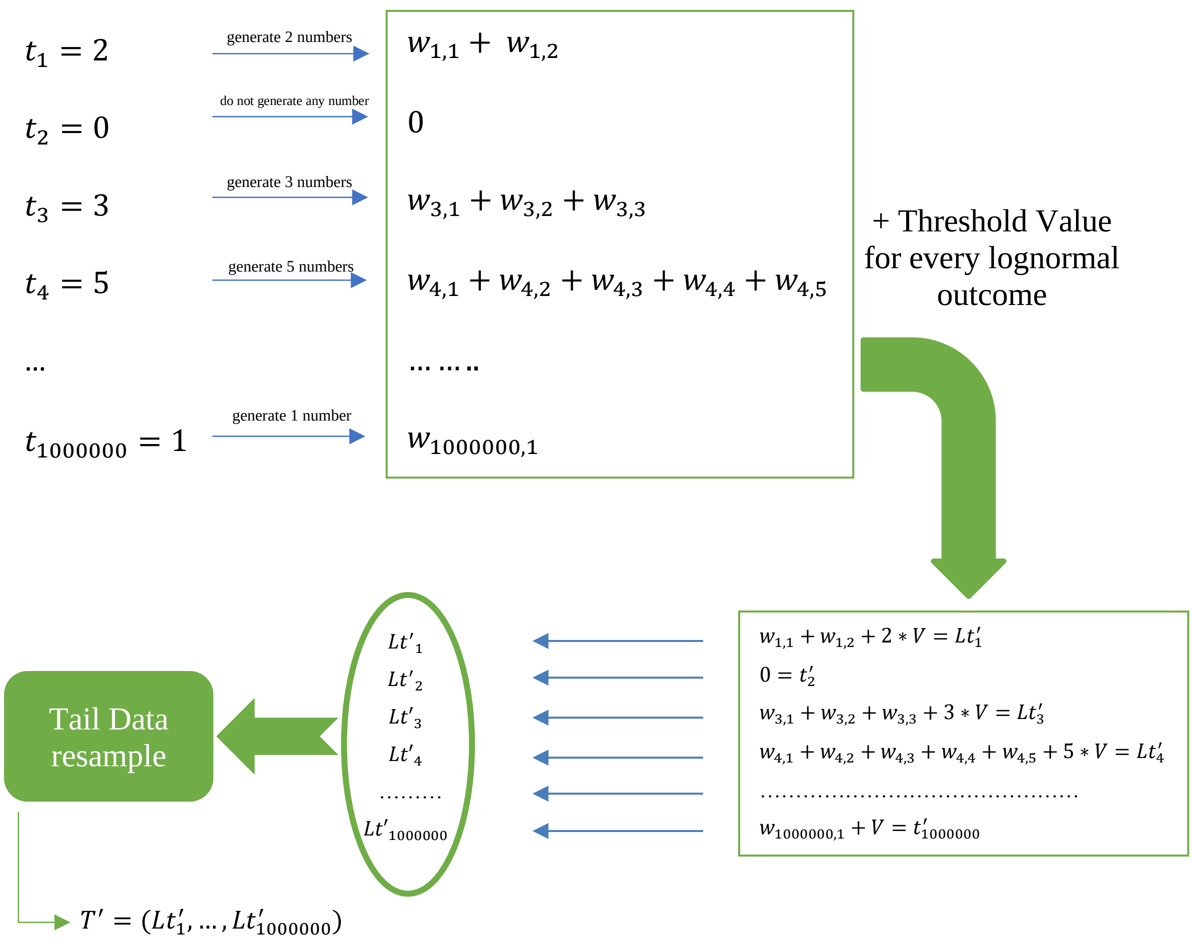

Figure 3.8. Algorithm to Simulate Tail Loss Distribution

Step 7: Sum all the values of the resample vectors and then obtain another array with the total data losses of each simulation made:

$$L = (L_1, L_2, \dots, L_{1000000}) = (Lb'_1, Lb'_2, \dots, Lb'_{1000000}) + (Lt'_1, Lt'_2, \dots, Lt'_{1000000}) \quad (8)$$

Step 8: Repeat all previous steps until all Risk Cells are addressed.

Finally, it follows a calibration of the Risk Cell losses to check the consistency of results, considering two types of estimation lines of the Minimum Capital Requirements (MCR) (Reynecke, 2018):

- **Undiversified**: The Risk Cell loss percentiles are summed as independent parcels, like the formula below:

$$MCR = L_p^1 + L_p^2 + \cdots + L_p^n = \sum_{i=1}^{n} L_p^i \tag{9}$$

$L_p^i - \ p^{th}\ quantile\ \ of\ the\ i^{th} Simulated\ Risk\ Cell\ loss\ (SRCL)$
$n - number\ of\ risk\ cells$

- **Fully Diversified**: The Risk Cells losses are summed, and then the percentile of that sum is estimated, as it follows:

$$MCR = p^{th}\ quantile\ of\ (SL_1, SL_2, \dots, SL_{1000000}) \tag{10}$$

$SL = SRCL_1 + SRCL_2 + \cdots + SRCL_n$
$SRCL_i - i^{th}\ Simulated\ Risk\ Cell\ Loss$
$n - number\ of\ risk\ cells$

Typically, banks adopt a more conservative position when estimating the capital losses, due to the uncertainty about its provisions, so the banks prefer a version that estimates higher amounts of capital (Anghelache & Olteanu, 2009). Since the Undiversified scenario previews a higher supplement than the other one, then the Undiversified is considered as the preferred approach to be used in OR management (Cooper, Piwcewicz, & Warren, 2014) and (Reynecke, 2018).

According to the CCR, "The operational risk measure shall capture potentially severe tail events, achieving a soundness standard comparable to a **99,9 % confidence interval over a one-year period**".

This means that the Minimum Capital Requirements for one year, according to this model, are the sum of the 99,9% quantiles of the Risk Cell losses simulated through Monte Carlo, as shown in equation (11):

$$MCR = \sum_{i=1}^{n} L_{99.9}^{(i)} \tag{11}$$

- $L_{99.9}^{(i)} -\ 99{,}9^{th}\ quantile\ of\ the\ simulated\ losses\ for\ risk\ cell\ i$

4. Application

As seen in the previous chapter, the results obtained with the model depend directly on the way the Risk Cells are built, since frequency and severity are simulated per Risk Cell. Basel II defines 56 Risk Cells in the quantification of the Operational Risk (seven types of risk affecting eight lines of business), but banks can use a different structure (Lambrigger, Shevchenko, & Wüthrich, 2007). So, as explained in Chapter 3, it was possible to use only the risk categories.

In our particular case study, it was necessary to aggregate the cells due to scarce data in most of them. Therefore, we performed two types of aggregation of the risk categories from the Capital Requirements Regulation, applying the Monte Carlo simulation to calculate the minimum capital requirements the bank must hedge against capital losses.

4.1. Database and Software

To develop the model, the algorithm was constructed in R language code. With the help of Excel, the database was imported from the bank and displayed in tables for the frequency, severity, and Monte Carlo results.

The data is available from 2008, but only years 2011-2021 (11 complete years) have been used, with the purpose of certifying that the losses belong to a period where the system of collection was sturdy, either in the method or the source, avoiding in this way anomalies in collected loss values.

Originally, the database contained 66 419 contracts that caused several incidents (with or without financial impact) and some of them were identified as risk sources. Since the OR model has influence in ICAAP conclusions, a decision was made to include only internal losses with financial impact, remaining 54 430 records of incidents. Also, the database consists of 65 fields, but only seven were used to build the model, namely:

- Registration Date, the date when the event was detected by the bank;
- Closing Date, the date of registration after the risk department verifies the loss occurrence and its regulatory classification;
- Loss Amount, corresponding to the Gross Loss Amount from equation (1) for both aggregation types;
- Recovered Amount, the capital directly recovered from the loss originated by the event (used for aggregation 1);
- Open Amount, the amount of loss potentially at risk (used for aggregation 2);
- Event Type, the risk category of the event (see Appendix 3)

- Group, to check if the institution to which the model is applied corresponds or not to the group of branches and subsidiaries.

The next two sections describe in detail the two aggregation procedures that have been followed, making clear the differences and similarities between them. One important aspect is that, after filtering all the information, only the Liquid Loss Amount, the date field, and the event type are required to construct the two variants of the model.

4.2. Aggregation 1

As already mentioned, only internal data with financial impact is considered. The date field is going to be the *registration date* and the Liquid Loss Amount is

$$\text{Liquid Loss Amount} = \max(\text{Loss Amount} - \text{Recovered Amount}, 0) \geq 250 \quad (12)$$

This resulted in 7549 cases, distributed among the seven categories described in article 324 of the CRR as shown in Figure 4.1.

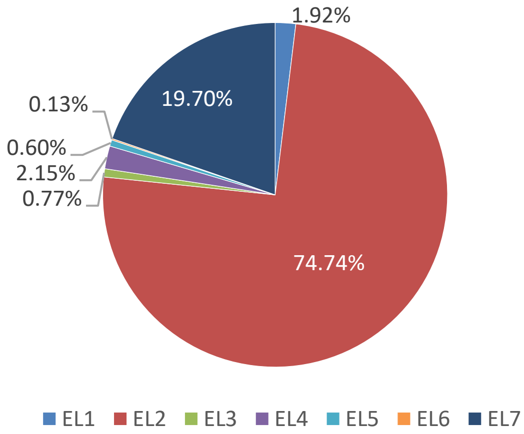

Figure 4.1. Distribution of Events Among the Seven Categories – Aggregation 1

In this variant, according to the bank's decision and the rules of aggregation described in section 3.2, it was decided to aggregate the categories into the following Risk Cells:

- **EL7 -** includes loss events classified as Execution, Delivery and Process Management;
- **ELgr -** includes events related with Internal Fraud (EL1); External Fraud (EL2); Employment Practices and Workplace Safety (EL3); Clients, Products and Business Practices (EL4); Damage to Physical Assets (EL5); Business Disruption and System Failures (EL6).

After mapping the event types for each contract according to Aggregation 1, the data is prepared for the frequency modelling. As explained in the methodology, for each Risk Cell their events were counted per Quarter, resulting in 44 Quarters (11 years x 4 quarters) that were labeled as in Figure 4.2.

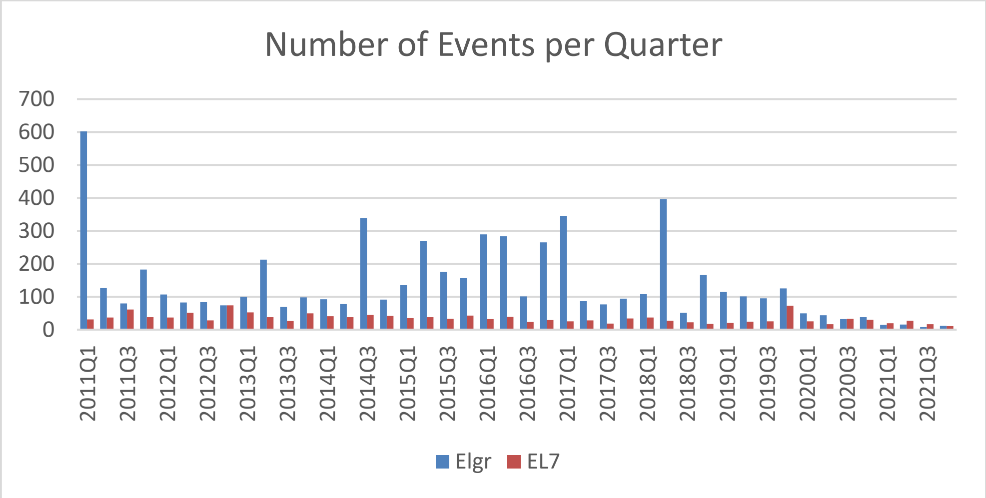

Figure 4.2. Frequency of Events per Quarter for Aggregation 1

From the data shown in Figure 4.2 (see Appendix 5), it was possible to estimate the frequency parameters:

Table 4.1. Frequency Parameters for Aggregation 1

Risk Cell Name	Mean	Variance	Critical Ratio	Lambda	Size	Mu
ELgr	137.77	14114.32	102.45	137.77	1.59	137.74
EL7	33.80	189.75	5.61	33.80	8.22	33.80

Observing the values in both Risk Cells, the Critical Ratio is much higher than 2, which means the Negative Binomial distribution will be selected to simulate the number of events. Looking into the Risk Cell ELgr, the huge mean and variance values are two remarkable aspects to notice, due to its massive representation in the dataset, more than 80% (80,30%) of total loss events and to the significant variability of the number of events from quarter to quarter (Figure 4.2).

Although the variance for the Risk Cell EL7 (20% of the losses) is lower, meaning the count of events does not fluctuate at so extraordinary levels, it has a Critical Ratio of 5,61, which is still high enough for the Negative Binomial to be chosen.

The next step was to decide which threshold value would be more suitable to define the Body and the Tail of the loss distribution, using the methodology explained in Chapter 3.

According to the bank, there is no external information about the Thresholds to be used. So, following the methodology from the severity model in 3.3.2, a percentile line must be chosen for each Risk Cell with sufficiently smooth parameters to guarantee an acceptable number of events in the Tail. Tables 4.2 and 4.3 provide an overview of the results for the two Risk Cells under analysis.

Table 4.2. Finding the Body-Tail threshold for Risk Cell EL7, Aggregation 1

	RiskCell	BodyWeig	NumberIn	NumberIn	Threshold	Threshold	Min_Tail	Q80_Tail	Q90_Tail	Max_Tail	FittedDistribution	Method	Paramete	Paramete
50%	7	0.000157	743	744	1268.9	0.5	1270.4	12556.4	31370.14	4782460	lognormal	maximum	8.520771	1.25577
50%	7	0.000157	743	744	1268.9	0.5	1270.4	12556.4	31370.14	4782460	gpd	maximum	0.714546	4929.6
60%	7	0.047141	595	892	1706.22	0.6	1706.7	18539.94	37318.44	4782460	lognormal	maximum	8.825567	1.22507
60%	7	0.047141	595	892	1706.22	0.6	1706.7	18539.94	37318.44	4782460	gpd	maximum	0.672351	6859.845
70%	7	0.099819	446	1041	2627.12	0.7	2636.4	26348.7	49227.1	4782460	lognormal	maximum	9.218298	1.175113
70%	7	0.099819	446	1041	2627.12	0.7	2636.4	26348.7	49227.1	4782460	gpd	maximum	0.605081	10550.28
80%	7	0.173681	298	1189	4856.28	0.8	4859.1	37318.44	87225.65	4782460	lognormal	maximum	9.749277	1.097371
80%	7	0.173681	298	1189	4856.28	0.8	4859.1	37318.44	87225.65	4782460	gpd	maximum	0.503198	18870.66
90%	7	0.3224	149	1338	12547.6	0.9	12574	87507.7	180112.7	4782460	lognormal	maximum	10.59632	0.944433
90%	7	0.3224	149	1338	12547.6	0.9	12574	87507.7	180112.7	4782460	gpd	maximum	0.352846	46136.69
91%	7	0.350121	133	1354	15000	0.91	15057.6	98595.3	2.00E+05	4782460	lognormal	maximum	10.72579	0.918023
91%	7	0.350121	133	1354	15000	0.91	15057.6	98595.3	2.00E+05	4782460	gpd	maximum	0.335433	52504.45
92%	7	0.379525	119	1368	18512.68	0.92	18812.5	100459	201210.9	4782460	lognormal	maximum	10.84437	0.898861
92%	7	0.379525	119	1368	18512.68	0.92	18812.5	100459	201210.9	4782460	gpd	maximum	0.322546	59157.78
93%	7	0.415724	105	1382	22792.94	0.93	22799	118819	208421.8	4782460	lognormal	maximum	10.96756	0.886712
93%	7	0.415724	105	1382	22792.94	0.93	22799	118819	208421.8	4782460	gpd	maximum	0.312803	66966.46
94%	7	0.461288	90	1397	26318.92	0.94	26348.7	142334	215464.5	4782460	lognormal	maximum	11.11186	0.87822
94%	7	0.461288	90	1397	26318.92	0.94	26348.7	142334	215464.5	4782460	gpd	maximum	0.305056	77322.99
95%	7	0.514685	75	1412	31340.51	0.95	31429.4	180112.7	231416.5	4782460	lognormal	maximum	11.28369	0.864718
95%	7	0.514685	75	1412	31340.51	0.95	31429.4	180112.7	231416.5	4782460	gpd	maximum	0.296912	91582.29
96%	7	0.57827	60	1427	37300.28	0.96	37500	201210.9	250235.5	4782460	lognormal	maximum	11.49659	0.841037
96%	7	0.57827	60	1427	37300.28	0.96	37500	201210.9	250235.5	4782460	gpd	maximum	0.290457	112136.9
97%	7	0.656363	45	1442	48904.59	0.97	51242.8	215494.5	270347.2	4782460	lognormal	maximum	11.78335	0.782314
97%	7	0.656363	45	1442	48904.59	0.97	51242.8	215494.5	270347.2	4782460	gpd	maximum	0.285537	145481
98%	7	0.779121	30	1457	87169.24	0.98	89200	252326.3	290686.1	4782460	lognormal	maximum	12.13402	0.732507
98%	7	0.779121	30	1457	87169.24	0.98	89200	252326.3	290686.1	4782460	gpd	maximum	0.299524	199663.9
99%	7	1	15	1472	178621.2	0.99	2.00E+05	295765.3	374559.5	4782460	lognormal	maximum	12.6047	0.768066
99%	7	1	15	1472	178621.2	0.99	2.00E+05	295765.3	374559.5	4782460	gpd	maximum	0.373289	308217.2

Table 4.3. Finding the Body-Tail threshold for Risk Cell ELgr, Aggregation 1

	RiskCell	BodyWeig	NumberIn	NumberIn	Threshold	Threshold	Min_Tail	Q80_Tail	Q90_Tail	Max_Tail	FittedDistribution	Method	Paramete	Paramete
50%	1	0.000201	3031	3031	823.8	0.5	823.9	2877.7	4786.6	51992640	lognormal	maximum	7.542498	0.906481
50%	1	0.000201	3031	3031	823.8	0.5	823.9	2877.7	4786.6	51992640	gpd	maximum	0.41271	1975.885
60%	1	0.12241	2425	3637	985.42	0.6	985.5	3393.14	5605.58	51992640	lognormal	maximum	7.724834	0.927393
60%	1	0.12241	2425	3637	985.42	0.6	985.5	3393.14	5605.58	51992640	gpd	maximum	0.43657	2341.483
70%	1	0.247109	1819	4243	1275.85	0.7	1276	4142.088	7014.84	51992640	lognormal	maximum	7.964779	0.956037
70%	1	0.247109	1819	4243	1275.85	0.7	1276	4142.088	7014.84	51992640	gpd	maximum	0.472106	2916.693
80%	1	0.393518	1213	4849	1796.56	0.8	1797.1	5605.58	10001.99	51992640	lognormal	maximum	8.28462	1.028974
80%	1	0.393518	1213	4849	1796.56	0.8	1797.1	5605.58	10001.99	51992640	gpd	maximum	0.542147	3900.958
90%	1	0.586088	607	5455	2877.52	0.9	2877.7	10001.99	26236.65	51992640	lognormal	maximum	8.86935	1.189543
90%	1	0.586088	607	5455	2877.52	0.9	2877.7	10001.99	26236.65	51992640	gpd	maximum	0.691783	6595.824
91%	1	0.610689	546	5516	3095.226	0.91	3096.5	11434	35678.2	51992640	lognormal	maximum	8.966782	1.215973
91%	1	0.610689	546	5516	3095.226	0.91	3096.5	11434	35678.2	51992640	gpd	maximum	0.716745	7195.382
92%	1	0.6373	485	5577	3392.684	0.92	3397.7	12507.56	44249.4	51992640	lognormal	maximum	9.07769	1.246746
92%	1	0.6373	485	5577	3392.684	0.92	3397.7	12507.56	44249.4	51992640	gpd	maximum	0.746972	7939.412
93%	1	0.665774	425	5637	3709.084	0.93	3712	16068.9	53210	51992640	lognormal	maximum	9.204148	1.282369
93%	1	0.665774	425	5637	3709.084	0.93	3712	16068.9	53210	51992640	gpd	maximum	0.781118	8886.09
94%	1	0.697199	364	5698	4140.56	0.94	4157.37	19241.18	69780.26	51992640	lognormal	maximum	9.360472	1.322743
94%	1	0.697199	364	5698	4140.56	0.94	4157.37	19241.18	69780.26	51992640	gpd	maximum	0.820977	10235.94
95%	1	0.732274	304	5758	4785.39	0.95	4786.6	26236.65	102317.8	51992640	lognormal	maximum	9.548478	1.371224
95%	1	0.732274	304	5758	4785.39	0.95	4786.6	26236.65	102317.8	51992640	gpd	maximum	0.86758	12112.24
96%	1	0.772642	243	5819	5605.428	0.96	5607.1	44249.4	135959.5	51992640	lognormal	maximum	9.801493	1.425724
96%	1	0.772642	243	5819	5605.428	0.96	5607.1	44249.4	135959.5	51992640	gpd	maximum	0.921282	15301.96
97%	1	0.820952	182	5880	7012.614	0.97	7074.2	70177.44	208590.3	51992640	lognormal	maximum	10.16052	1.482997
97%	1	0.820952	182	5880	7012.614	0.97	7074.2	70177.44	208590.3	51992640	gpd	maximum	0.974515	21558.06
98%	1	0.883947	122	5940	10001.94	0.98	10002.49	135959.5	271808.8	51992640	lognormal	maximum	10.72076	1.524393
98%	1	0.883947	122	5940	10001.94	0.98	10002.49	135959.5	271808.8	51992640	gpd	maximum	0.994764	38064.36
99%	1	1	61	6001	26236.13	0.99	26267.9	273686.4	783257.9	51992640	lognormal	maximum	11.84404	1.431529
99%	1	1	61	6001	26236.13	0.99	26267.9	273686.4	783257.9	51992640	gpd	maximum	0.884181	123986.7

The percentile 98% was chosen for ELgr and percentile 93% was chosen for EL7, returning a threshold of 10001,94€ and 22792,94€ respectively for the two Risk Cells. As so, it was possible to proceed with the estimation of the parameters for Lognormal distribution, required to simulate the loss amounts exceeding the thresholds for the two Risk Cells. Results are in Table 4.4.

Table 4.4. Severity parameters for Aggregation 1

Risk Cell	Threshold Quantile	Threshold	Fitted Distribution	Parameter 1	Parameter 2	Bodyweight
ELgr	0.98	10001.94	Lognormal	10.7208	1.5244	0.8839
EL7	0.93	22792.94	Lognormal	10.9676	0.8867	0.4157

In other words, this means that 88,39% of the outcomes for ELgr must be simulated from the body part (41,57% for EL7), and the remaining 11,61% (58.43%) must be simulated from the tail part, applying the algorithms described in Chapter 3.

Tables 4.5-4.8 show the results of Monte Carlo method (in Euros). A sequence of 16 tests have been performed (4 tests for each of $k = 1000, k = 10000, k = 100000$ and $k = 1000000$ iterations) to assess the model's efficiency and to verify if there is a relation between the number of simulations and the accuracy in the returned outputs.

Table 4.5. Monte Carlo Results After 1000 Iterations, Aggregation 1

Names	RiskCell	ExpectedLoss	90%	99%	99.90%	99.99%
EL7	7	10878602.20	23785442.41	39692465.82	56120480.15	65381413.92
ELgr	1	8814593.06	13307675.79	19087886.30	23639751.15	23722389.29
Undiv.	Undiv.	19693195.26	37093118.20	58780352.12	79760231.30	89103803.20
Fully. Div.	Fully. Div.	19693195.26	33453931.45	49482904.87	69338211.84	78708676.43

Names	RiskCell	ExpectedLoss	90%	99%	99.90%	99.99%
EL7	7	10338890.85	21717481.35	43606030.11	54612361.26	68584013.49
ELgr	1	9019106.033	13610620.79	17875148.48	21210574.35	24465101.65
Undiv.	Undiv.	19357996.89	35328102.15	61481178.59	75822935.61	93049115.14
Fully. Div.	Fully. Div.	19357996.89	31270615.49	54044453.55	66365083.82	74323043.95

Names	RiskCell	ExpectedLoss	90%	99%	99.90%	99.99%
EL7	7	10130652.24	22274600.03	38873285.81	56526154.50	63496670.15
ELgr	1	8730797.05	13038976.88	18004884.73	19528709.28	19797520.92
Undiv.	Undiv.	18861449.29	35313576.90	56878170.54	76054863.79	83294191.07
Fully. Div.	Fully. Div.	18861449.29	31370240.41	50041987.29	65019651.38	71544742.00

Names	RiskCell	ExpectedLoss	90%	99%	99.90%	99.99%
EL7	7	10409356.49	22291213.18	40835558.06	62930945.39	71050875.08
ELgr	1	8619068.46	13052482.13	18357142.53	20668694.75	22936650.76
Undiv.	Undiv.	19028424.94	35343695.30	59192700.59	83599640.14	93987525.84
Fully. Div.	Fully. Div.	19028424.94	31376924.61	49850562.88	76180515.99	76637465.34

Table 4.6. Monte Carlo Results After 10000 Iterations, Aggregation 1

Names	RiskCell	ExpectedLoss	90%	99%	99.90%	99.99%
EL7	7	10546515.74	22543601.53	42762106.59	61882855.54	76079352.84
ELgr	1	8708105.31	13115476.85	18028973.42	22205621.45	25519438.13
Undiv.	Undiv.	19254621.05	35659078.38	60791080.02	84088476.99	101598790.97
Fully. Div.	Fully. Div.	19254621.05	32156950.10	52198347.76	70045738.01	85498610.62

Names	RiskCell	ExpectedLoss	90%	99%	99.90%	99.99%
EL7	7	10563147.83	22612463.73	42134878.12	63480711.23	79989486.58
ELgr	1	8742962.13	13184170.72	18009417.04	22702121.30	24823809.47
Undiv.	Undiv.	19306109.96	35796634.45	60144295.16	86182832.53	104813296.05
Fully. Div.	Fully. Div.	19306109.96	31936338.53	51842123.28	69373378.76	89558209.98

Names	RiskCell	ExpectedLoss	90%	99%	99.90%	99.99%
EL7	7	10571566.90	22563064.32	42753531.52	58527947.33	81542970.41
ELgr	1	8687460.23	13068039.05	17795020.52	21845163.20	25437591.45
Undiv.	Undiv.	19259027.14	35631103.36	60548552.05	80373110.54	106980561.85
Fully. Div.	Fully. Div.	19259027.14	31870598.69	51358332.71	71334009.92	90458680.20

Names	RiskCell	ExpectedLoss	90%	99%	99.90%	99.99%
EL7	7	10564070.68	22431488.14	42276212.7	57609269.37	67669528.55
ELgr	1	8698051.54	13047666.05	17901891.7	22471452.09	26314299.01
Undiv.	Undiv.	19262122.22	35479154.19	60178104.4	80080721.45	93983827.56
Fully. Div.	Fully. Div.	19262122.22	32095011.01	51343019.82	65223716.91	74422458.98

Table 4.7. Monte Carlo Results After 100000 Iterations, Aggregation 1

Names	RiskCell	ExpectedLoss	90%	99%	99.90%	99.99%
EL7	7	10537034.48	22563143.05	42233561.77	61908927.35	87261569.14
ELgr	1	8715609.48	13083588.77	18035955.90	22603776.92	26148064.73
Undiv.	Undiv.	19252643.96	35646731.83	60269517.67	84512704.27	113409633.87
Fully. Div.	Fully. Div.	19252643.96	31996530.25	51662648.14	71711232.05	93949929.94

Names	RiskCell	ExpectedLoss	90%	99%	99.90%	99.99%
EL7	7	10474634.49	22461606.16	41680030.29	61171228.65	83621560.06
ELgr	1	8718161.14	13117688.82	17994451.23	22111189.25	26292473.74
Undiv.	Undiv.	19192795.62	35579294.98	59674481.52	83282417.90	109914033.80
Fully. Div.	Fully. Div.	19192795.62	31942563.17	50934611.52	71131406.11	92277548.32

Names	RiskCell	ExpectedLoss	90%	99%	99.90%	99.99%
EL7	7	10537487.94	22559977.49	42056938.64	63284651.00	81707646.15
ELgr	1	8708964.19	13075498.92	18005038.45	22253052.50	25317281.95
Undiv.	Undiv.	19246452.13	35635476.41	60061977.09	85537703.50	107024928.11
Fully. Div.	Fully. Div.	19246452.13	31925092.15	51371609.97	72897120.24	94234143.80

Names	RiskCell	ExpectedLoss	90%	99%	99.90%	99.99%
EL7	7	10480998.64	22349131.84	41864038.65	62288344.39	96532925.81
ELgr	1	8730944.52	13103742.24	17976366.85	22312555.79	27150516.64
Undiv.	Undiv.	19211943.16	35452874.08	59840405.50	84600900.19	123683442.44
Fully. Div.	Fully. Div.	19211943.16	31643761.62	51292793.84	70943918.52	104958375.71

Table 4.8. Monte Carlo Results After 1000000 Iterations, Aggregation 1

Names	RiskCell	ExpectedLoss	90%	99%	99.90%	99.99%
EL7	7	10481912.20	22402431.38	41794723.52	61498409.96	86545229.22
ELgr	1	8717433.83	13110877.26	18047572.23	22337891.10	26292251.41
Undiv.	Undiv.	19199346.03	35513308.63	59842295.75	83836301.06	112837480.62
Fully. Div.	Fully. Div.	19199346.03	31801779.41	51250772.68	71038304.13	96212569.99

Names	RiskCell	ExpectedLoss	90%	99%	99.90%	99.99%
EL7	7	10488180.78	22434878.35	41868449.06	61919029.45	84383921.67
Elgr	1	8725024.84	13127118.19	18024949.62	22277754.43	26209923.40
Undiv.	Undiv.	19213205.62	35561996.55	59893398.68	84196783.88	110593845.06
Fully. Div.	Fully. Div.	19213205.62	31844077.78	51400898.97	71500264.80	92911413.66

Names	RiskCell	ExpectedLoss	90%	99%	99.90%	99.99%
EL7	7	10482448.78	22409003.95	41696209.83	61689034.76	86777648.86
ELgr	1	8722805.59	13115872.36	18053682.89	22277945.49	25934338.41
Undiv.	Undiv.	19205254.37	35524876.32	59749892.72	83966980.25	112711987.27
Fully. Div.	Fully. Div.	19205254.37	31805922.38	51165017.97	71100047.40	96069587.11

Names	RiskCell	ExpectedLoss	90%	99%	99.90%	99.99%
EL7	7	10483646.02	22506551.49	42784574.92	60902312.28	94000633.95
ELgr	1	8734803.33	13137383.19	17475780.44	21969306.37	24883444.35
Undiv.	Undiv.	19218449.35	35643934.67	60260355.35	82871618.65	118884078.30
Fully. Div.	Fully. Div.	19218449.35	32018869.64	51387043.97	71041444.39	101873002.77

To analyze further the results, the whole procedure was repeated assuming a different aggregation, Aggregation 2. The purpose is to seek confirmation of the results and make conclusions more reliable.

4.3. Aggregation 2

Likewise the first aggregation, only internal information with financial impact was used, but with three relevant differences:

- The date field is the closing date (accounting date);
- The Liquid Loss Amount formula is:

$$\text{Liquid Loss Amount} = \max(\text{Loss Amount} + \text{Open Amount}, 0) \geq 250 \qquad (13)$$

- Only losses occurred in Portuguese entities were included.

After applying these rules 8313 cases remained, representing the new variant of the Operational Risk model, which includes more events. This is due to the Liquid Loss Amount formula, which adds a strictly positive variable (the Open Amount) to the Loss amount, surpassing easier the 250€ amount. This allows more events to be included, and the number of events that belong to non-Portuguese institutions is not enough to maintain the same number of events as in Aggregation 1.

Figure 4.3 displays how these events are distributed among the seven categories described in article 324 of the CRR.

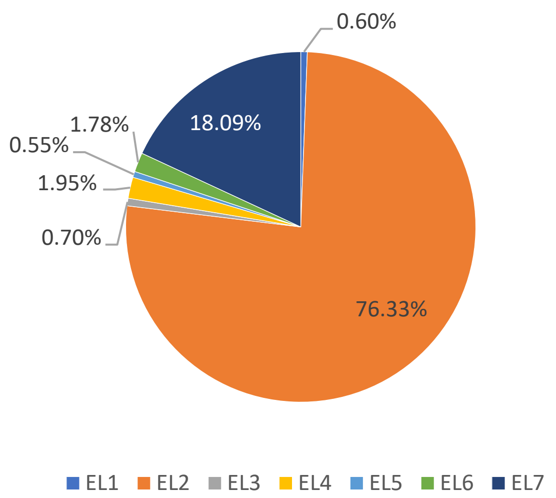

Figure 4.3. Distribution of Events Among the Seven Categories – Aggregation 2

This time, the bank decided to perform a new aggregation, considering again two Risk Cells, but as two groups of categories instead of isolating one category as an individual Risk Cell from the remaining ones. Like in the first aggregation, the EL2 and EL7 categories present the major parts of the dataset (see Figures 4.2-4.3), being EL2 76% of the total losses registered and EL7 18%. So, the bank decided to join the less representative categories into these two: EL5 is aggregated into EL2 for being the smallest one, constituting 46 events of the dataset (**ELgr1**), while EL1, EL3, EL4 and EL6 are joint to EL7, forming the **ELgr2**.

From now on, all the steps will replicate what has been done for Aggregation 1. Figure 4.4 shows the number of occurrences per Risk Cell and per quarter.

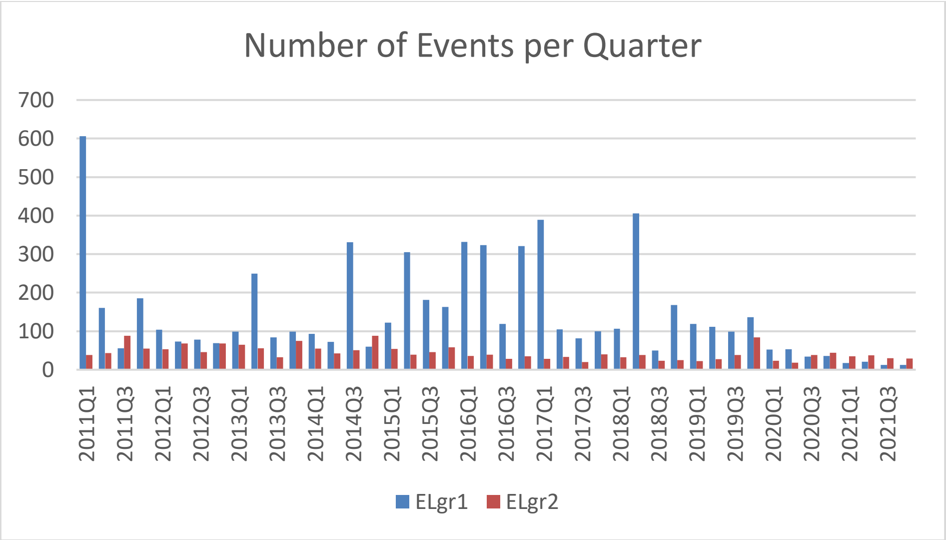

Figure 4.4. Frequency of Events per Quarter for Aggregation 2

From this data (see Appendix 5), it was possible to calculate the frequency parameters (time measured in quarters):

Table 4.9. Frequency Parameters for Aggregation 2

Risk Cell Name	Mean	Variance	Critical Ratio	Lambda	Size	Mu
ELgr1	145.25	16116.42	110.96	145.25	1.56	145.28
ELgr2	43.68	324.22	7.42	43.68	7.66	43.68

Looking at these results, it is possible to see similarities with what was observed in Aggregation 1, in the sense of having again an extremely high ratio for ELgr1 (76,88% of the full dataset) and both Critical Ratios being greater than 2. The Negative Binomial distribution is the model chosen for the frequency variable of the two Risk Cells.

For this simulation, it was not necessary to perform a fitting table to define the Body-Tail threshold for the loss distributions, because the bank required the thresholds to be 10000€ for ELgr1 and 100000€ for ELgr2. We used then these amounts to estimate the parameters of the Lognormal distribution and to calculate the corresponding bodyweights, in order to proceed with the Monte Carlo simulation.

Table 4.10. Severity Parameters for Aggregation 2

Risk Cell	Threshold	Fitted Distribution	Parameter 1	Parameter 2	Bodyweight
ELgr1	10000	Lognormal	10.0472	1.1525	0.9835
ELgr2	100000	Lognormal	12.5862	1.1163	0.6543

Now 98,34% of the outcomes for ELgr1 must be simulated from the body part (65,43% for ELgr2), and the remaining 1,66% (34.57%) from the tail part.

Tables 4.11-4.14 below show the results of Monte Carlo method in Euros. Again, a sequence of 16 tests was performed (four tests for each of each k, $k = 1000, k = 10000, k = 100000$ and $k = 1000000$ simulations for the same reasons given in Section 4.2, before presenting the results for Aggregation 1.

Table 4.11. Monte Carlo Results After 1000 Iterations for Aggregation 2

Names	RiskCell	ExpectedLoss	90%	99%	99.90%	99.99%
ELgr1	1	1256925.38	2616882.24	4819388.06	6165430.50	6480800.59
ELgr2	2	38219177.87	60827320.50	85209444.04	97690324.63	101816397.49
Undiv.	Undiv.	39476103.25	63444202.74	90028832.09	103855755.12	108297198.08
Fully. Div.	Fully. Div.	39476103.25	62070142.82	86556636.32	98635736.68	102160235.50

Names	RiskCell	ExpectedLoss	90%	99%	99.90%	99.99%
ELgr1	1	1230536.17	2604955.71	5022522.61	6069437.90	7160521.40
ELgr2	2	39242071.32	60123722.48	84096182.74	111190608.29	120402706.56
Undiv.	Undiv.	40472607.49	62728678.19	89118705.35	117260046.19	127563227.97
Fully. Div.	Fully. Div.	40472607.49	61561601.56	86160639.93	111638184.77	121359497.74

Names	RiskCell	ExpectedLoss	90%	99%	99.90%	99.99%
ELgr1	1	1223003.05	2664706.01	4987879.68	6582082.42	6890553.99
ELgr2	2	39760531.25	62107801.16	86007528.04	116693559.60	118774597.12
Undiv.	Undiv.	40983534.31	64772507.17	90995407.72	123275642.02	125665151.11
Fully. Div.	Fully. Div.	40983534.31	63593039.05	86349599.59	119134969.86	119802682.25

Names	RiskCell	ExpectedLoss	90%	99%	99.90%	99.99%
ELgr1	1	1233697.97	2671536.44	4631802.11	6139870.42	6359232.81
ELgr2	2	40318614.71	61305192.74	89279136.90	107398339.39	140264938.33
Undiv.	Undiv.	41552312.69	63976729.18	93910939.01	113538209.81	146624171.14
Fully. Div.	Fully. Div.	41552312.69	62429348.26	90251602.29	112215405.46	141712676.47

Table 4.12. Monte Carlo Results After 10000 Iterations for Aggregation 2

Names	RiskCell	ExpectedLoss	90%	99%	99.90%	99.99%
ELgr1	1	1218024.65	2578140.32	4697676.59	6734575.29	8192448.35
ELgr2	2	39676209.94	62055661.65	87252123.54	109654444.73	121071295.26
Undiv.	Undiv.	40894234.59	64633801.96	91949800.14	116389020.01	129263743.61
Fully. Div.	Fully. Div.	40894234.59	63202122.13	89061592.62	110135291.28	122565780.79

Names	RiskCell	ExpectedLoss	90%	99%	99.90%	99.99%
ELgr1	1	1213517.356	2570986.894	4655280.976	6413494.165	7349261.909
ELgr2	2	39649455.09	61636399.95	87204213.3	111179928.5	129474889.7
Undiv.	Undiv.	40862972.44	64207386.84	91859494.28	117593422.6	136824151.6
Fully. Div.	Fully. Div.	40862972.44	62810414.5	88281768.57	111905839.6	130335635.9

Names	RiskCell	ExpectedLoss	90%	99%	99.90%	99.99%
ELgr1	1	1187383.507	2488879.126	4499485.639	6286158.215	7435982.487
ELgr2	2	39884735.27	62125821.73	86744620.1	106340341.1	124573113.2
Undiv.	Undiv.	41072118.78	64614700.85	91244105.74	112626499.3	132009095.7
Fully. Div.	Fully. Div.	41072118.78	63425484.97	88138016.02	108011198.5	125039816.3

Names	RiskCell	ExpectedLoss	90%	99%	99.90%	99.99%
ELgr1	1	1206858.369	2526679.789	4658788.412	6552875.25	7931285.963
ELgr2	2	39873403.57	62031327.16	87950768.77	107910896.9	125102783.7
Undiv.	Undiv.	41080261.94	64558006.95	92609557.18	114463772.2	133034069.7
Fully. Div.	Fully. Div.	41080261.94	63378404.2	89197990.63	109065608.6	126583199.7

Table 4.13. Monte Carlo Results After 100000 Iterations for Aggregation 2

Names	RiskCell	ExpectedLoss	90%	99%	99.90%	99.99%
ELgr1	1	1215214.167	2563377.331	4629100.657	6782397.155	8792220.849
ELgr2	2	39643546.36	61926990.89	87774826.64	112078722.5	134995346
Undiv.	Undiv.	40858760.53	64490368.23	92403927.29	118861119.7	143787566.9
Fully. Div.	Fully. Div.	40858760.53	63134561.94	89077742.55	112907313.4	135512678.5

Names	RiskCell	ExpectedLoss	90%	99%	99.90%	99.99%
ELgr1	1	1216460.11	2552116.33	4695213.76	6708049.65	9006941.19
ELgr2	2	39612523.29	61822057.06	87635993.88	110265382.59	130054109.75
Undiv.	Undiv.	40828983.39	64374173.39	92331207.63	116973432.24	139061050.93
Fully. Div.	Fully. Div.	40828983.39	63121644.46	88966070.22	111481380.27	131818300.29

Names	RiskCell	ExpectedLoss	90%	99%	99.90%	99.99%
ELgr1	1	1212832.14	2569655.31	4659355.32	6758767.64	8421236.41
ELgr2	2	39611185.67	61650483.76	87444962.52	112258184.30	132634053.83
Undiv.	Undiv.	40824017.81	64220139.07	92104317.84	119016951.94	141055290.24
Fully. Div.	Fully. Div.	40824017.81	62915971.03	88855655.86	113191699.40	134179722.25

Names	RiskCell	ExpectedLoss	90%	99%	99.90%	99.99%
ELgr1	1	1211912.21	2557628.60	4687841.10	6635200.38	8956879.69
ELgr2	2	39517367.86	61583715.92	86987751.38	111428762.54	128813128.81
Undiv.	Undiv.	40729280.08	64141344.52	91675592.48	118063962.92	137770008.50
Fully. Div.	Fully. Div.	40729280.08	62822164.30	88286956.71	112516751.12	130549139.51

Table 4.14. Monte Carlo Results After 1000000 Iterations for Aggregation 2

Names	RiskCell	ExpectedLoss	90%	99%	99.90%	99.99%
ELgr1	1	1213985.10	2562608.24	4676505.25	6734852.62	8992202.89
ELgr2	2	39597625.30	61678786.59	87244804.13	109787319.43	131601662.33
Undiv.	Undiv.	40811610.40	64241394.83	91921309.38	116522172.05	140593865.22
Fully. Div.	Fully. Div.	40811610.40	62939143.28	88537433.42	111192848.99	133041710.19

Names	RiskCell	ExpectedLoss	90%	99%	99.90%	99.99%
ELgr1	1	1212091.13	2559894.36	4673140.00	6703391.19	8664566.56
ELgr2	2	39577469.05	61600141.79	87204742.01	110031663.19	133119792.78
Undiv.	Undiv.	40789560.19	64160036.15	91877882.01	116735054.37	141784359.35
Fully. Div.	Fully. Div.	40789560.19	62840802.66	88437502.10	111223843.63	134454699.15

Names	RiskCell	ExpectedLoss	90%	99%	99.90%	99.99%
ELgr1	1	1212386.88	2558998.81	4653638.88	6717485.01	8765168.26
ELgr2	2	39562203.32	61632822.40	87230314.64	109521774.05	129660939.42
Undiv.	Undiv.	40774590.20	64191821.21	91883953.53	116239259.06	138426107.68
Fully. Div.	Fully. Div.	40774590.20	62875467.90	88523048.48	110587490.00	131227305.70

Names	RiskCell	ExpectedLoss	90%	99%	99.90%	99.99%
ELgr1	1	1213428.80	2554723.80	4675956.41	6704225.49	8696909.29
ELgr2	2	39609700.68	61689043.22	87477045.63	110294528.10	131369636.98
Undiv.	Undiv.	40823129.48	64243767.02	92153002.04	116998753.59	140066546.27
Fully. Div.	Fully. Div.	40823129.48	62941290.97	88721887.62	111527945.11	132819033.63

4.4. Results and Discussion

Since this model is stochastic, the results vary whenever it is made a simulation, which justifies the need to perform four tests for each input number. This allows to study the sensibility of oscillation on the results every time a simulation is running.

Analyzing the given outputs carefully, for both models it is visible that the expected loss and the losses associated to each quantile generated by Monte Carlo tend to fluctuate less with the increase of the number of simulations.

Looking into Aggregation 1, the expected losses oscillate between 18,8 and 19,7 million Euros (M€), which results a difference of around 831000€ each time the Monte Carlo replicates $k = 1000$ simulations, but if it is set $k = 1000000$ iterations, the results lean towards 19,2 M€, with a maximum difference around 19000€. Although in the case of $k = 10000$ and $k = 100000$ simulations differ similarly in the expected loss, it is evident that the quantiles are more unstable for the first input number than the second. For example, the 99,9% percentile goes from 80,1 to 86,2 M€ for $k = 10000$ simulations, whereas for $k = 100000$ it waves between 83,3 and 85,5 M€.

Moving to Aggregation 2, the expected loss varies between 40,77 and 40,83 M€, returning then a difference of almost 50000€ when $k = 1000000$, while in the case of $k = 1000$ simulations the expected loss differs between 39,4 and 41,6 M€, which corresponds to an imprecision of around 2,1 M€.

Focusing only on the results for $k = 1000000$ and assuming the capital regulatory is the mean of the results displayed in the four tables (column of quantile 99,9% and row the Undiversified, see end of Section 3.4), from the first model it is estimated a capital regulatory of 83,7 M€ to the bank with all the restrictions, aggregations and orders demanded to ensure they have enough capital to perform all the operations and protect against market randomness. The second model predicts 116,6 M€ of event losses to the 99,9% quantile according to the bank's requests accomplished.

Furthermore, it is evident that the expected losses are closer between the two groups in the first model than between those in the second model. Also, in both models, the Risk Cells that contained less categories (less diverse) generated more capital losses than the more diverse Risk Cells.

In this case, the first model EL7 generated around 10,48M of loss capital and ELgr 8,72M, being this last Risk Cell a more diversified group that covers six event types, while in the second model ELgr1 generates losses of amount 1,21M and ELgr2 of amount 39,6M. ELgr1, which has a major slice of the data by event type, is little diversified, containing only events related with Damage to Physical Assets and External Fraud.

This analysis may raise many questions about the outcomes, but throughout the report some variables that could modify the values of frequency and severity parameters were identified, and therefore the results through Monte Carlo model.

There is a 'suspicion' about the Liquid Loss Amount formula, because in the first model the recovered amount is deducted while in the second the open amount is added, and since both formulas present only non-negative values, the Liquid Loss Amount values in general will be higher in Aggregation 2 - which may explain why the total expected loss and the quantiles for Undiversified and Fully Diversified Scenarios present higher values than the ones in Aggregation 1.

Another aspect to point out is in the severity table, specifically in the selected Threshold value for each Risk Cell, because it determines how many events are considered in the Tail a key aspect to quantify the simulated loss amounts. Those allow to estimate the bodyweights, the Mu (the mean), and the Sigma (standard deviation) parameters through maximum likelihood estimation. In Aggregation 1, the highest threshold corresponds to EL7, and it was 22792,94€ while in Aggregation 2 was 100000€ for ELgr2, which is 4,39 times larger. This means that in Aggregation 2, the tails for the loss distribution are thicker than in Aggregation 1, because the Threshold is multiplied by the number of events generated for the Tail distribution[10], and it is enough that one of these variables increases for a higher loss value to appear.

Although the controlling variables can modify the results, the initial dataset is not mutable, which implies that if the OR model is applied to other datasets with the same structure and column names, it is impossible to change the original column values including the event types of each occurrence. For this study approach, 94,4% of the occurrences in the filtered dataset belong either to the ‘External Fraud’ or ‘Execution, Delivery & Process Management’ loss event types, for both models, remaining then the other 5,6% to the other five categories, which means there are not many choices of aggregations to study the frequency and severity besides the ones made.

[10] 6th step of the Monte Carlo simulation in chapter 3.4

5. Conclusion

This framework demonstrates an empirical approach to measure the capital risk through advanced methods, with the purpose of obtaining Minimum Capital Requirements for effects of ICAAP decision making.

It is a model that can be explored and manually adjusted, for example in the Liquid Loss Amount formula, in the Risk Cells and in the Thresholds, creating opportunities to analyze some of them.

Since this model does not follow all the quantitative standards from the CRR (article 322), it may be seen as incomplete.

Nevertheless, I think it is an interesting model that can be computed in programming languages. Additionally, it is not very common to find a project that directly applies mathematical knowledge and programming skills to test and experiment in the financial sector. The essay explores a hypothesis possibly helpful to the world of financial mathematics, thus presenting an opportunity to exhibit the academic knowledge for a real-life project.

Finally, this academic internship in KPMG gave me tools and knowledge useful to my future career, it was an excellent experience to progress in the professional world.

References

Anghelache, G. V., & Olteanu, A. C. (2009). The Operational Risk-Minimum Capital Requirements. *Theoretical & Applied Economics*, 16(7).

Anghelache, G., Olteanu, A. C., & Radu, A. N. (2010). Operational Risk Measurement. *European Research Studies*.

Audit Services | IFRS 9. (2017). Retrieved from PwC: https://www.pwc.com/gx/en/audit-services/ifrs/publications/ifrs-9/ifrs-9-understanding-the-basics.pdf

Bantleon, U., d'Arcy, A., Eulerich, M., Hucke, A., Pedell, B., & Ratzinger-Sakel, N. V. (2021). Coordination challenges in implementing the three lines of defense. *International Journal of Auditing*, 25(1), 59-74.

Bermúdez, M. R. (2015). Operational Risk modeling: A Review.

Brent, R. P. (1971). An algorithm with guaranteed convergence for finding a zero of a function. *The computer journal*, 14(4), 422-425.

Cooper, B., Piwcewicz, B., & Warren, N. (2014). *Operational risk modelling: how far have we progressed?* Sydney: Financial Services Forum, Actuaries Institute Australia.

Crowley, S. (2012). Maximum Likelihood Estimation of the Negative Binomial Distribution.

Davis, J., & Haubenstock, M. (2002). Building Effective Indicators to Monitor Operational Risk. *The RMA Journal*, 84(8), 40-43.

De Jonghe, O., Dewachter, H., & Ongena, S. (2020). Bank Capital (Requirements) and Credit Supply: Evidence from Pillar 2 Decisions. *Journal of Corporate Finance*, Volume 60.

Dutta, K., & Perry, J. (2006). *A tale of tails: An empirical analysis of loss distribution models for estimating operational risk capital.* Boston.

European Bank Authority. (2015). *Final Draft RTS on AMA Assessment for Operational Risk.*

Farid, J. A. (2010). *Internal Capital Adequacy Assessment Process (ICAAP)-Overview & Core Concepts.*

Fazio, G., & Modica, M. (2012). *Pareto or log-normal? A recursive-truncation approach to the distribution of (all) cities.*

Final Draft RTS on AMA Assessment for Operational Risk. (2015, June 3). *European Bank Authority*. Retrieved from European Bank Authority.

FSI | IFRS 9 and expected loss provisioning - Executive Summary. (n.d.). Retrieved from Bank of International Settlements: https://www.bis.org/fsi/fsisummaries/ifrs9.pdf

Hadley v Baxendale - 1854. (2013). Retrieved from LawTeacher: https://www.lawteacher.net/cases/hadley-v-baxendale.php?vref=1

Jöhnemark, A. (2012). *Modeling Operational Risk.*

Korn, R., Korn, E., & Kroisandt, G. (2010). *Monte Carlo Methods and Models in Finance and Insurance.* London: Chapman & Hall/ CRC Financial Mathematics Series.

KPMG. (2021). Pilar I: Solvência, TRIM e Risco Operacional.

KPMG Portugal. (n.d.). *Serviços | Advisory.* Retrieved from https://home.kpmg/pt/pt/home/services/advisory.html

KPMG Portugal. (n.d.). *Serviços | Audit and Assurance.* Retrieved from https://home.kpmg/pt/pt/home/services/audit.html

KPMG Portugal. (n.d.). *Serviços | Tax.* Retrieved from https://home.kpmg/pt/pt/home/services/tax.html

Lambrigger, D., Shevchenko, P., & Wüthrich, M. (2007). The Quantification of Operational Risk using Internal Data, Relevant External Data and Expert Opinions. *Journal of Operational Risk*, 3-27.

Lantz, B. (2013). *Machine Learning with R: Machine Learning with R: Learn how to use R to apply powerful machine learning methods and gain an insight into real world applications.* Birmingham: Packt Publishing Ltd.

Luo, T., Zhou, L., Falzon, J., Cheng, Y., Christianson, G., Wu, Y., & Habchi, A. (2021). Assessing Physical Climate Risks for the European Bank for Reconstruction and Development's Power Generation Project Investment Portfolio. *World Resources Institute*. Retrieved from World Resources Institute.

Opdyke, J., & Cavallo, A. (2012). Estimating operational risk capital: the challenges of truncation, the hazards of MLE, and the promise of robust statistics. *The Journal of Operational Risk, Forthcoming.*

OPE Calculation of RWA for Operational Risk. (1999). In *Basel Committee on Banking Supervision.* Bank for International Settlements.

Pavlovich, M. (2022, February 16). *Log-normal Distribution - A simple explanation.* Retrieved from towards data science: https://towardsdatascience.com/log-normal-distribution-a-simple-explanation-7605864fb67c

Reynecke, W. (2018). *Quantitative response to the operational risk problems of external data scaling and dependence structure optimization for capital diversification.*

Risk Wiki | Aggregate Monte Carlo. (n.d.). Retrieved from Vose Software: https://www.vosesoftware.com/riskwiki/VoseAggregateMonteCarlo.php

Rochet, J. (2004). Rebalancing the three pillars of Basel II. *Economic Policy Review*.

Scenarios Portal. (n.d.). Retrieved from NGFS: https://www.ngfs.net/ngfs-scenarios-portal/

Septiany, R., Setiawaty, B., & Purnaba, I. P. (2020). The Use of Monte Carlo Method to Model the Aggregate Loss Distribution. *Al-Jabar: Jurnal Pendidikan Matematika,* 11(1), 179-190.

Sobre a KPMG. (n.d.). Retrieved from KPMG Portugal: https://home.kpmg/pt/pt/home/about.html

The European Parliament. (2013). Regulation (EU) No 575. *Council of the European Union* (pp. 194-200). Official Journal of the European Union.

Weinberg, J. (2013). *The Great Recession and Its Aftermath.* Retrieved from https://www.federalreservehistory.org/essays/great-recession-and-its-aftermath

What are the Big 4 Consulting Firms? (n.d.). Retrieved from CFA Journal: https://www.cfajournal.org/big-4-consulting-firms/

What's the impact on expected credit losses? (n.d.). Retrieved from KPMG: https://home.kpmg/xx/en/home/insights/2021/06/climatechange-ifrs9-ecl.html

Appendix

Appendix 1 – Internship Plan Chronogram

	March	*April*	*May*	*June*	*July*	*August*
Training	X					
Key point analysis		X				
Metrics development			X	X	X	
Results analysis					X	X
Report preparation						X

Source: Author’s elaboration

Appendix 2 – Process and Internal Data Quantitative Standards

2. The standards relating to Process are the following:

a) an institution shall calculate its own funds requirement as comprising both expected loss and unexpected loss, unless expected loss is adequately captured in its internal business practices. The operational risk measure shall capture potentially severe tail events, achieving a soundness standard comparable to a 99,9 % confidence interval over a one year period;

b) an institution's operational risk measurement system shall include the use of internal data, external data, scenario analysis and factors reflecting the business environment and internal control systems as set out in paragraphs 3 to 6. An institution shall have in place a well documented approach for weighting the use of these four elements in its overall operational risk measurement system;

c) an institution's risk measurement system shall capture the major drivers of risk affecting the shape of the tail of the estimated distribution of losses;

d) an institution may recognise correlations in operational risk losses across individual operational risk estimates only where its systems for measuring correlations are sound, implemented with integrity, and take into account the uncertainty surrounding any such correlation estimates, particularly in periods of stress. An institution shall validate its correlation assumptions using appropriate quantitative and qualitative techniques;

e) an institution's risk measurement system shall be internally consistent and shall avoid the multiple counting of qualitative assessments or risk mitigation techniques recognised in other areas of this Regulation.

3. The standards relating to Internal Data are the following:

a) an institution shall base its internally generated operational risk measures on a minimum historical observation period of five years. When an institution first moves to an Advanced Measurement Approach, it may use a three-year historical observation period;

b) an institution shall be able to map their historical internal loss data into the business lines defined in Article 317 and into the event types defined in Article 324, and to provide these data to competent authorities upon request. In exceptional circumstances, an institution may allocate loss events which affect the entire institution to an additional business line "corporate items". An institution shall have in place documented, objective criteria for allocating losses to the specified business lines and event types. An institution shall record the operational risk losses that are related to credit risk and that the institution has historically included in the internal credit risk databases in the operational risk databases and shall identify them separately. Such losses shall not be subject to the operational risk charge, provided that the institution is required to continue to treat them as credit risk for the purposes of calculating own funds requirements. An institution shall include operational risk losses that are related to market risks in the scope of the own funds requirement for operational risk;

c) an institution's internal loss data shall be comprehensive in that it captures all material activities and exposures from all appropriate sub-systems and geographic locations. An institution shall be able to justify that any excluded activities or exposures, both individually and in combination, would not have a material impact on the overall risk estimates. An institution shall define appropriate minimum loss thresholds for internal loss data collection;

d) aside from information on gross loss amounts, an institution shall collect information about the date of the loss event, any recoveries of gross loss amounts, as well as descriptive information about the drivers or causes of the loss event;

e) an institution shall have in place specific criteria for assigning loss data arising from a loss event in a centralised function or an activity that spans more than one business line, as well as from related loss events over time;

f) an institution shall have in place documented procedures for assessing the on-going relevance of historical loss data, including those situations in which judgement overrides, scaling, or other adjustments may be used, to what extent they may be used and who is authorised to make such decisions

Source: EU Regulation 575/2013, article 322(2-3)

Appendix 3 – Risk Categories and their respective Labels

Event type	Risk Cell Label
Internal Fraud	**EL1**
External Fraud	**EL2**
Employment Practices and Workplace Safety	**EL3**
Clients, Products and Business Practices	**EL4**
Damage to Physical Assets	**EL5**
Business Disruption and System Failures	**EL6**
Execution, Delivery & Process Management	**EL7**

Source: Article 324 of the European Regulation 575/2013

Appendix 4 – Paragraph 667 from the Basel II

"Given the continuing evolution of analytical approaches for operational risk, the Committee is not specifying the approach or distributional assumptions used to generate the operational risk measure for regulatory capital purposes. However, a bank must be able to demonstrate that its approach captures potentially severe 'tail' loss events. Whatever approach is used, a bank must demonstrate that its operational risk measure meets a soundness standard comparable to that of the internal ratings-based approach for credit risk (i.e. comparable to a one year holding period and a 99.9th percentile confidence interval)."

Source: Basel Committee on Banking Supervision

Appendix 5 – Tables with the Number of Events per Quarter for Aggregations 1 and 2

	Elgr	EL7
2011Q1	602	31
2011Q2	126	37
2011Q3	80	61
2011Q4	182	38
2012Q1	107	37
2012Q2	82	51
2012Q3	83	28
2012Q4	74	74
2013Q1	100	52
2013Q2	213	38
2013Q3	69	26
2013Q4	98	49
2014Q1	92	41
2014Q2	78	38
2014Q3	339	45
2014Q4	91	42
2015Q1	135	35
2015Q2	270	38
2015Q3	176	33
2015Q4	156	43
2016Q1	289	32
2016Q2	283	39
2016Q3	101	23
2016Q4	265	29
2017Q1	346	25
2017Q2	86	28
2017Q3	77	18
2017Q4	94	34
2018Q1	108	37
2018Q2	396	27
2018Q3	51	22
2018Q4	166	17
2019Q1	114	20
2019Q2	101	24
2019Q3	95	25
2019Q4	125	73
2020Q1	49	25
2020Q2	44	16
2020Q3	32	33
2020Q4	38	30
2021Q1	14	19
2021Q2	15	27
2021Q3	8	16
2021Q4	12	11

	ELgr1	ELgr2
2011Q1	606	38
2011Q2	160	43
2011Q3	56	88
2011Q4	185	55
2012Q1	104	53
2012Q2	73	68
2012Q3	78	46
2012Q4	69	68
2013Q1	99	65
2013Q2	249	56
2013Q3	84	32
2013Q4	99	75
2014Q1	93	55
2014Q2	72	42
2014Q3	331	51
2014Q4	60	88
2015Q1	122	54
2015Q2	305	39
2015Q3	181	46
2015Q4	163	58
2016Q1	332	36
2016Q2	323	39
2016Q3	119	28
2016Q4	321	35
2017Q1	389	28
2017Q2	105	33
2017Q3	81	20
2017Q4	100	40
2018Q1	106	32
2018Q2	406	38
2018Q3	50	23
2018Q4	168	25
2019Q1	119	22
2019Q2	111	27
2019Q3	99	38
2019Q4	136	84
2020Q1	52	23
2020Q2	53	18
2020Q3	34	38
2020Q4	36	44
2021Q1	17	35
2021Q2	21	37
2021Q3	12	30
2021Q4	12	29

Source: Author’s elaboration