Master Actuariat de Dauphine

Mémoire présenté devant l'Université Paris Dauphine
pour l'obtention du diplôme du Master Actuariat
et l'admission à l'Institut des Actuaires

le ________________________

Par : Shin Woong KANG

Titre: Quantification of Operational Risk within Insurance company

Confidentialité : ☐ NON ☑ OUI (Durée : ☐ 1 an ☑ 2 ans)

Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus

Membre présent du jury de l'Institut des Actuaires :

Signature :

Entreprise : AXA GRM

Nom : KANG Shin Woong

Signature :

Directeur de mémoire en entreprise :

Nom : CHANSIN Olivier

Signature :

Membres présents du jury du Master Actuariat de Dauphine :

***Autorisation de publication et de mise en ligne sur un site de diffusion de documents actuariels** (après expiration de l'éventuel délai de confidentialité)*

Signature du responsable entreprise :

Secrétariat :

Bibliothèque :

Signature du candidat :

Université Paris-Dauphine, Place du Maréchal de Lattre de Tassigny, 75775 PARIS Cedex 16

Quantification of Operational Risk within an Insurance company

Université Paris-Dauphine

Shin Woong KANG

09/10/2015

Acknowledgements

I would like to take this opportunity to give my thanks to all the people that I have met during my internship.

First, of all, I want to thank Olivier CHANSIN for being my tutor. I really appreciated his patience with me, his advices and for mentoring my work through the internship.

I also want to thank all the team members of the Operational Risk team of AXA GRM – Céline SAMAIN, Jean-Baptiste PETIT and Mai SAYACHACK – who were always there when I needed help and who considered me as a team member.

A big thank to all the members of the GRM, Julie LEROUVILLOIS, Estèphe ARNAUD, Alexandre SAN MARTIN… It was a pleasure having you around during my lunchtime or my breaks. I am sure I forgot a lot of people.

My experience at the GRM was a good experience; I had the chance to have really nice, dynamic, skilled and smart people around me.

Abstract

The definition of Operational Risk is a loss that results from inadequate or failed internal processes or losses that result from people, systems or from external events. In the banking industry, Basel II has defined regulatory requirements for the quantification and computation of the Operational Risk capital charge. For the insurance industry, Solvency II introduced the concept of Operational Risk within Insurance Company. With the heterogeneity and the lack of data, modeling and quantifying the operational risk can be challenging. The purpose of this master thesis is to present the different approach for the measurement of the operational risk with the associated mathematical and actuarial concepts with a focus on the AXA Group model: the Scenario Based Approach (SBA). We will in this master thesis compare the approach of AXA's internal model with another approach, the Loss Distribution Approach (LDA).

Keywords: Operational Risk, SCR, Solvency II, frequency-severity, SBA, aggregation approach, correlation matrices, copula, Monte Carlo, LDA

Résumé

Le Risque Opérationnel se définit comme le risque d'avoir une perte résultant de la défaillance ou de l'inadéquation d'une procédure interne, de son personnel, des systèmes internes ou à des risques externes. Dans l'industrie financière, Bâle II a introduit des exigences réglementaires pour la quantification et le calcul d'une charge en capital de Risque Opérationnel. Avec l'hétérogénéité et le manque de de données, modéliser et quantifier le Risque Opérationnel peut s'avérer difficile. L'objectif de ce mémoire est de présenter les différentes approches pour mesurer le Risque Opérationnel avec les concepts mathématiques et actuariels associés avec un focus sur le modèle d'AXA Group : Scenario Based Approach (SBA), ou encore l'approche par scénario. On va pour cela comparer le modèle interne d'AXA, le SBA, avec une autre approche, la LDA (Loss Distribution Approach), ou l'approche par estimation des distributions de pertes.

<u>Mots clefs :</u> Risque Opérationnel, SCR, Solvabilité II, fréquence-sévérité, SBA, approche par agrégation, matrices de corrélations, copule, Monte Carlo, LDA

Executive Summary

Operational Risk is a risk that exists in every company, not only on the Insurance industry. There are a lot of famous operational risk losses: Madoff's Ponzi scheme is one of them. Due to the significant impact of operational losses on the company's solvency, the Solvency II directive, in line with Basel II, has defined requirements for insurance and reinsurance companies to have sufficient capital and an effective risk management framework for Operational Risk as for the other risks that they face in their activities (e.g. insurance risks, financial risks, etc…) There are two way of quantifying its operational risk are proposed:

- Standard Formula
- Internal Model

Given the scarcity and heterogeneity of the data, modeling operational risk can be a really challenging task and can be time-consuming as well as costly for the company. One of the alternatives to the use of such internal data is to use one of the alternatives approaches introduced with the Solvency II directive: the Scenario Based Approach (SBA). We compensate scarcity and heterogeneity of the data with the use of expert judgments. The steps of the quantification process will be described from the frequency-severity approach to the aggregation of all the quantified risks.

The first chapter will describe the AXA Group, and will describe the concept of operational risk with the Solvency II definition and some well-known examples of operational losses. The framework of the operational risk is presented as well. The chapter 2 will be presenting the two way of quantifying the operational risk capital charge. Each Insurance company can have its own methodology to quantify their operational risk, and more specifically, to compute the operational risk capital charge. The choice is let to the Insurance Company between the Standard formula and an internal model. If the internal model is chosen, the Insurance Company has the choice of the approach, the most used approach in the Insurance Industry are:

- Loss Distribution Approach (LDA)
- Scenario Based Approach (SBA)
- Bayesian Approach

We will see that the Basel II regulations have largely inspired the Solvency II regulations. In fact, Basel II introduced the Advanced Measurement Approach (AMA) for the operational risk quantification, which is largely used in the Insurance industry for the same purpose. The

Chapters 3 and 4 will focus on AXA's approach for its internal model, the SBA. The steps of the quantification from the risk identification to the scenario calibration and the capital computation will be presented. The main principle of the SBA is to identify, assess and measure the most important operational risks faced by Insurance company. It is also a forward-looking approach that adequately represents the risk profile of the company. The Monte Carlo approach to simulate a total loss distribution will be presented, as well as a study that demonstrates the convergence and stability of the Monte Carlo approach. As operational risks often result on low frequency high severity losses, a study on the frequency threshold is introduced to find a threshold for the frequency parameter that ensures us a positive capital charge, which is the Value-at-Risk at a 99.5% level with a one-year horizon.

Chapter 5 introduces the concept of the diversification of the risks, as an Insurance company has to hold a capital charge for Operational Risk. The simple sum of all the risks is not adequate, because all operational risks are not 100% dependent between them. Indeed, it is very intuitive to conceive that all operational risks do not occur at the same time and in every entity of AXA. It is expected that a fire incident in Italy would not be correlated to the flood in Paris. Therefore, we cannot expect operational risks to be completely independent from each other. Two approaches to diversify the operational risks are presented:

- Variance – Covariance
- Copula

The two approaches have both advantages and drawbacks. We will see that the second approach will be used, as it better catches the dependencies between the distribution tails. But in order to apply the Copula approach, the correlation matrix, which is built from the correlation between the risks or entities, have some needed properties. Due to the fact that in a SBA, the correlation matrices are built upon expert judgments, they do not always have all these needed properties. One of them is the positive definiteness of the correlation matrix, which is essential during the implementation of the Copula approach. For instance, two algorithms that transform a non-Positive Definite (PD) matrix to a PD one are presented. As an example of correlations used for the diversification effect, there are two studies on correlations in this chapter:

- Correlation between Operational Risk categories
- Geographical correlation (or correlation between entities)

The two types of correlations have been studied with the internal data. The studies' aim is to determine whether or not the correlation levels are low. Different types of correlation measure are introduced such as the Pearson's correlation coefficient or the Kendall's rank correlation.

The last chapter, Chapter 6 aims to quantify the capital charge of different operational risk categories with a LDA. The different categories were tested with non-parametric statistical tools such as the Kolmogorov-Smirnov or Anderson-Darling adequacy tests. Then, with the calibrated frequency and severity distribution with a LDA, we compute the capital charges of the operational risk categories. The differences between the capital charges obtained with a LDA, and with a SBA (AXA's approach) will be highlighted and discussed to find an alternative to enhance AXA's current internal model.

Note de synthèse

Le Risque Opérationnel est un risque qui existe dans toutes les entreprises, et pas seulement dans l'industrie de l'assurance. On peut citer de nombreux exemples de pertes opérationnelles comme l'escroquerie de type « Ponzi » de Madoff. En raison de l'impact important des récentes pertes opérationnelles sur la solvabilité d'une entreprise, la directive Solvabilité II, dans la lignée de Bâle II, a défini des exigences pour les compagnies d'assurance et de réassurance. Celles ci doivent détenir un montant de capital adéquat et avoir un cadre efficace de gestion des risques pour le Risque Opérationnel et les autres risques que celles-ci encourent dans leur activité (ex : risque assurantiel, risque financier,...). Il existe deux façons de quantifier le risque opérationnel sont proposées dans la directive :

- la formule standard
- Modèle interne

Compte tenu de la rareté et de l'hétérogénéité des données, la modélisation du risque opérationnel peut être une tâche qui s'avère être difficile, peut être chronophage et coûteux pour l'entreprise. Une des alternatives à l'utilisation de ces données interne est d'utiliser l'une des approches alternatives introduites avec la directive Solvabilité II: le Scenario Based Approach (SBA). Nous compensons la rareté et l'hétérogénéité des données avec l'utilisation de jugements d'experts. Les étapes du processus de quantification seront décrites depuis l'approche fréquence sévérité jusqu'à l'agrégation de tous les risques quantifiés.

Le premier chapitre décrit le Groupe AXA, et décrit les concepts du risque opérationnel avec la définition de Solvabilité 2 et quelques exemples célèbres de pertes opérationnelles. Le cadre de gestion du risque opérationnel y est aussi présenté. Le chapitre 2 présentera les deux manières de quantifier la charge de capital de risque opérationnel. Chaque compagnie d'assurance peut adopter une méthodologie propre à elle-même pour quantifier leur risque opérationnel, et plus spécifiquement, pour calculer la charge de capital de risque opérationnel. Le choix est laissé à la compagnie d'assurance entre la formule standard et un modèle interne. Si le modèle interne est choisi, la compagnie d'assurance a le choix de l'approche. Les approches les plus utilisées dans l'industrie assurantielle sont :

- Loss Distribution Approach (LDA), approche par estimation des distributions de perte
- Scenario Based Approach (SBA), approche par scénario
- Approche Bayésienne

Nous allons voir que la directive Solvabilité 2 a été largement inspirée des réglementations de Bâle. En effet, Bâle II a introduit l'Advanced Measurement Approach (AMA), pour la quantification du risque opérationnel, qui est largement utilisé dans l'industrie de l'assurance. Les chapitres 3 et 4 porteront sur l'approche d'AXA pour son modèle interne, la SBA. Les étapes de la quantification à partir de l'identification des risques jusqu'à la calibration des scénarios et aussi jusqu'au calcul de la charge en capital. Le principe essentiel de la SBA est d'identifier, d'évaluer et de mesurer les risques opérationnels les plus importants auxquels la compagnie d'assurance fait face. C'est également une approche prospective qui représente convenablement le profil de risque de l'entreprise. L'approche de Monte Carlo pour simuler une distribution de perte totale sera présentée, ainsi que d'une étude qui démontre la convergence et la stabilité de l'approche de Monte Carlo. Comme les risques opérationnels résultent souvent de pertes dites de « basse fréquence haute sévérité », une étude sur le seuil de fréquence est introduite pour trouver un seuil pour le paramètre de fréquence qui nous assure une charge de capital positive, qui est la Value-at-Risk à un niveau de 99,5% à horizon un an.

Le Chapitre 5 introduit le concept de la diversification des risques. Toute compagnie d'assurance se doit d'avoir une charge de capital pour le risque opérationnel. La simple somme de tous les risques ne convient pas, parce que tous les risques opérationnels ne sont pas 100% liés entre eux. En effet, il est très intuitif de concevoir que tous les risques opérationnels ne se produisent pas en même temps et dans toutes les entités d'AXA. Il est évident qu'un incendie en Italie ne serait pas corrélé aux crues de la Seine à Paris. Par conséquent, nous ne pouvons pas attendre que les risques opérationnels soient totalement indépendants les uns des autres. Deux approches pour diversifier les risques opérationnels sont présentées:

• Variance - covariance

• Copule

Les deux approches ont des avantages et des inconvénients. Nous allons voir que la deuxième approche sera retenue par AXA, car elle prend mieux en compte les dépendances entre les queues de distribution. Mais afin d'appliquer l'approche par copule, la matrice de corrélation, qui est construit à partir de la corrélation entre les risques ou les entités, ont des propriétés nécessaires et indispensables. En raison du fait que, dans une SBA, les matrices de corrélation sont construites à partir de jugements d'experts, elles n'ont pas toujours toutes les propriétés nécessaires. L'une d'eux est que la matrice de corrélation doit être définie positive, cette est essentiel, surtout lors de la mise en œuvre de l'approche par copule. Nous avons ainsi deux algorithmes qui transforment une matrice non-définie positive (DP) à un une matrice DP. Pour illustrer avec un exemple l'effet de diversification, deux études sur les corrélations sont présentées dans ce chapitre:

• Corrélation géographique (ou de corrélation entre les entités)

• Corrélation entre catégories de risques opérationnels

Les deux types de corrélations ont été étudiés avec les données internes. Le but des études est de déterminer si oui ou non le niveau de ces corrélations est bas. Différents types de mesure de corrélation sont introduits tels que le coefficient de corrélation de Pearson ou le coefficient de corrélation de rang de Kendall.

Le dernier chapitre, le chapitre 6 vise à quantifier la charge de capital des différentes catégories de risque opérationnel avec une LDA. Les différentes catégories ont été testées avec des outils statistiques non-paramétriques tels que le test d'adéquation d'Anderson-Darling ou de Kolmogorov-Smirnov. Enfin, avec les distributions de fréquence et sévérité calibrées avec une LDA, nous calculons les charges de capital de ces catégories de risques opérationnels. Les écarts entre les charges de capital obtenus avec une LDA, et avec une SBA (l'approche d'AXA) seront mis en évidence et seront le sujet d'une discussion afin de trouver une alternative pour améliorer le modèle interne d'AXA.

Table of contents

Chapter 1 Introduction

1. AXA Group

This section aims to describe the structure of AXA Group Risk Management (GRM). The risk profile of the GRM will be studied in this thesis. It is largely inspired by AXA's internal communication information available in the intranet. AXA chose to gather its central functions within an economic interest grouping (GIE). In France, a GIE is a grouping that enables its members to share some of their activities in order to develop, improve and increase the results of these activities while keeping their own individuality at the same time. The Group Risk Management is within the GIE of AXA.

a) Group Risk Management

Risk management within Insurance companies is about identifying and selecting the risks. The Group Risk Management's role is to create a secure framework encouraging effectively managed underwriting, guaranteeing protection for the company over the long term.

The main missions and activities of the GRM:

- Creating a secure risk framework:

By clarifying decision-making in order to ensure better risk selection. For this, AXA has defined a series of standards, such as the Product Approval Process (PAP), covering all new products before their market release, or Risk Appetite, defining the limits within which the Group would like to operate.

- Protecting the company over the long term:

By testing AXA's capacity to overcome all types of major crises. For this AXA uses a series of stress scenarios to assess our capacity to withstand rare and extreme conditions with multiple impacts.

- Implementing the Solvency II project:

By driving the process for approval with the regulator and implementing our internal model for Short Term Economic Capital to guide decision-making.

- Developing the risk culture:

Building awareness among operational staff on the importance of a good risk management allows an enhancement of the risk culture. For this, AXA works to develop a strong risk culture through training and communication.

The Group Operational Risk Team belongs to the GRM.

b) Group Operational Risk Team

This team aims to ensure the good implementation of the framework across the Group to systematically identify, measure, mitigate, report and monitor the most important and significant Operational Risks that AXA may face through the implementation of risk assessment and mitigation processes and a loss data collection. The team covers all insurances operating entities, banking activities, asset managers… It coordinates a network of local operational risk teams, which conduct processes of local ownership regarding operational risk and has for mission to:

- Coordinate the entities
- Organize trainings
- Develop the risk culture in line with Solvency II Directives

Alongside with the above responsibilities, the team is also in charge of defining and enhancing the modelling approach of Operational Risk within AXA's Internal Model and to provide support to entities in the correct application of the methodology.

2. What is Operational Risk?

Operational Risk is a risk that exists in every company. It can be defined as the risk of having losses related to an error of process, a process failure, human error, IT system disruption or external event. In the Basel II regulations for the banking industry, Operational Risk is "the risk of a change in value caused by the fact that actual losses, incurred for inadequate or failed internal processes, people and systems, or from external events (including legal risk), differ from the expected losses". With the new Solvency II Directive, a similar definition of Operational Risk is applied within insurances companies.

Operational risks don't only exist in the banking or insurance industry. They are risks that can be found in every company. Some examples are miss click that could lead to a wrong/late product delivery, a case of internal fraud from a tied agent or a flood in Paris. Depending on the gravity of the flood, a significant number of employees will have to be relocated, thus creating additional costs and affecting the productivity of the company during the event.

According to the duration of the event, an economic impact similar to a recession could affect the whole company. Additional costs with a productivity downturn could be fatal for the company and lead it to bankruptcy. Operational risk can be very expensive but is fundamental to better understand the risk profile of a company.

<u>Famous example of operational risk losses:</u>

Volkswagen (2015)

> In September 2015, the US Environmental Protection Agency discovered that a "defeat device", a programme was installed in the engine software of more than 11 million cars. The device aim is to detect if the car is being driven under test conditions, reducing the NO_X[1]. As an example, the cars on American roads were emitting up to 30-40 times more toxic fumes than permitted. The scandal, rapidly named Volkswagate on Twitter and the news has already done damages to the firm. The CEO Martin Winterkorn announced his resignation the 23rd of September 2015. Also, Volkswagen is subject to an $18bn fine and expects to spend over $7.3bn on fixes and compensation.

Barings (1995)

> The Britain's oldest bank, Barings went to bankruptcy in 1995. It was the result of Nick Lesson's actions, who was the head derivative trader and head of the settlement operations in Singapore. Usually, two different people do these jobs. It allowed Lesson to speculate on the Asia market, and report losses as gains to Barings headquarters. The fraudulent actions of Lesson combined with the disastrous economic impact of the earthquake of Kobe turned the market against him. The losses amounted to GBP827m, twice the available trading capital of Barings and the bankruptcy itself cost another GBP100m. Barings was finally declared insolvent inly three days after the losses discovery and ING, a banking and insurance company bought it for GBP1. As a consequence of his actions, Lesson was sentenced to six years and half of prison and to a fine of GBP70k.

There are many other cases of big operational losses: Madoff's Ponzi scheme or Allen Stanford "massive Ponzi scheme".

[1] Generic term for the mono-nitrogen oxides: nitric oxide (NO) and nitrogen dioxide (NO_2). In large cities, the amount of NO_X emitted into the atmosphere as air pollution can be significant.

3. Solvency II

In the insurance industry, the operational risk framework is defined in the Solvency II Directive. Its aim is to strengthen the solvency capital requirements so that all the commitments of the insurance company can be met. These directives are largely inspired and similar to the ones in the Basel Committee.

The Solvency II Directive, or the Directive 2009/138/EC was adopted in November 2009. In 2011 the implementing measures started and the Directive is to enter in force in January 2016. Solvency II introduces a new and harmonized European wide regulatory framework. Solvency II has a tree pillars approach:

- Pillar 1: Quantitative Capital Requirements

In the first pillar, quantitative rules were defined for the computation of an Insurance company own fund. Two levels of own funds, the MCR[2] and SCR[3] are set to ensure the company's solvency and the protection of the clients from a possible default. Regulator can have an intervention if the own funds are judged insufficient. All the computations are on based on the Value-at-Risk with a 99.5% confidence level over a one-year horizon.

- Pillar 2: Qualitative Supervisory Review

The pillar 2 sets out the requirements for insurers' governance and risk management. It is a new supervisory system and new key functions are defined: compliance, internal audit, internal control, and risk management. All insurers must undertake the ORSA[4], which is an internal risk assessment process or an auto-evaluation of the risk management.

- Pillar 3: Market Discipline

The last pillar is focused on the Insurers' disclosure and transparency to the different stakeholders. New reporting processes are introduced to harmonize reporting requirements all across Europe.

With the Solvency II regulations, a new topic is introduced, the Operational Risk measurement. Pillar 1 corresponds to the quantification of the capital charge required. The management of the operational risk is introduced in Pillar II with the ERM[5]. We can read in the Solvency II directives that:

> "The Solvency Capital Requirement shall cover at least the following risks: (a) non-life underwriting risk, (b) life underwriting risk, […], (f) operational risk. Operational risk

[2] MCR = Minimum Capital Requirement
[3] SCR = Solvency Capital Requirement
[4] Own Risk and Solvency Assessment
[5] Enterprise Risk Management

[…] shall include legal risks, and exclude risks arising from strategic decisions, as well as reputation risks.”

4. AXA’s Operational Risk Framework

a) Definition

AXA’s definition for operational risk is aligned with the Solvency II one:

“Operational risk is the risk of loss arising from inadequate or failed internal processes, or from personnel and systems, or from external events”. It is also detailed that:

- Failure or inadequacy may result from both internal and external causes
- This definition includes legal risk
- It excludes reputation risk and risk arising from strategic decisions

In order to best define its Operational Risk, AXA highlights the main differences between Operational Risk and the “classic risks” such as insurance and financial risks. In these two classic risks, more risk means more expected returns. It is not the case in Operational Risk. More risk only means more losses.

b) AXA’s Operational Risks’ cartography

Segmentation of such heterogeneous set of risks as operational risk constitutes a crucial assessment process. A common approach starts by using the operational risk categories inspired from the Basel II framework:

IF	Internal Fraud
EF	External Fraud
EPWS	Employment Practices and Workplace Safety
CPBP	Clients, Products and Business Practices
DPA	Damage to Physical Assets
BDSF	Business Disruption and System Failure
EDPM	Execution, Delivery and Process Management

Table 1 Basel II operational risk categories

These seven categories are divided into 17 sub-categories as follow:

Internal Fraud	Internal Fraud & intentional unauthorized activity
External Fraud	External Fraud & system security fraud (hacking)
Employment Practices and Workplace Safety	Employee relations, diversity & discrimination
	Safe environment (e.g. pandemic)
	Loss of key staff / talent management
Clients, Products and Business Practices	Suitability, disclosure & breach of fiduciary duty (e.g. misselling, aggressive sales, misleading marketing materials,…) & advisory activities
	Improper business or market practices leading to breach of laws regulations (Antitrust, unlicensed activity, money laundering, breach of insurance laws)
	Product flaws (product defects & model errors)
Damage to Physical Assets	Disasters & other events (e.g. natural disasters or man-made losses)
Business Disruption and System Failure	Breach of information security confidentiality, integrity risk, data privacy)
	System disruption (unavailability, under-investment in IT or telecommunication)
Execution, Delivery and Process Management	Transaction, capture, execution & maintenance (erroneous payments, …)
	Failures in u/w or claims (non-intentional) processes
	Public reporting & disclosure risk (including financial reporting, rating issues & IR)
	Project management failure
	Internal vendors misperformance or failure, incl. distribution misperformance (tied agents)
	External vendors misperfomance or failure, incl. distribution misperformance (brokers)

Table 2 Operational risk sub-categories

Chapter 2: Modeling of Operational Risk in Industry Practice

Insurers have two ways of quantifying and modeling their operational risks: the standard approach or the internal model approach. The standard approach is the simplest. The Operational Risk SCR is computed as a percentage of their earned premiums and technical provisions. The internal model approach is to measure the risk that corresponds to the actual situation of the insurer. Its advantages are to create a risk culture and awareness to the risk managers and actuaries. The internal model can be a key tool for risk management and can allow a reduction of the capital charge.

With the new Solvency II Directive, every Insurance Company has to allocate a capital charge for Operational Risk. There are two possible ways for its quantification:

- the Standard Formula: the Operational Risk capital charge or SCR is computed using a factor based approach
- internal model: the company measure and quantify its Operational Risk according to its risk profile

The two approaches have their advantages and drawbacks. On one side, the Standard Formula is easy to compute as it is a close formula, but the Operational Risk SCR in this case does not correspond to the risk profile of the company. On the other side, the internal model has the advantage to compute an operational risk SCR that corresponds more to the risk profile of the company. The drawback is that measuring and quantifying operational risks is not trivial. Also, the model has to be validated by the local regulator.

1. Standard formula

Many firms today use the standard formula for computing their operational risk capital charge. The formula that will be described is the one issued by CEIOPS in QIS5 (July 2010).

The inputs for the Operational Risk module are the following:

- $TP_{life} =$ Life insurance obligations. For the purpose if this calibration, technical provisions should not include the risk margin, should be without deduction of recoverable from reinsurance contracts and special purpose vehicles
- $TP_{life-ul} =$ Life insurance obligations for life insurance obligations where the investment risk is borne by the policyholders. For the purpose of this calculation, technical provisions should not include risk margin, should be

without deduction of recoverable from reinsurance contracts and special purpose vehicles

- TP_{nl} = Total non-life insurance obligations excluding obligations under on-life contracts which are similar to obligations, including annuities. For the purpose of this calculation, technical provisions should not include the risk margin and should be without deduction of recoverable from reinsurance contract and special purpose vehicles
- $pEarn_{life}$ =Earned premium during the 12 months prior to the previous 12 month for life insurance obligations, without deducting the premium ceded to reinsurance
- $pEarn_{life-ul}$ = Earned premium during the 12 months prior to the previous 12 month for life insurance obligations where the investment risk is borne by the policyholders, without deducting the premium ceded to reinsurance
- $pEarn_{nl}$ =Earned premium during the 12 months prior to the previous 12 month for non-life insurance obligations, without deducting the premium ceded to reinsurance
- $Earn_{life}$ = Earned premium during the previous 12 month for life insurance obligations, without deducting the premium ceded to reinsurance
- $Earn_{life-ul}$ = Earned premium during the previous 12 month for life insurance obligations where the investment risk is borne by the policyholders, without deducting the premium ceded to reinsurance
- $Earn_{nl}$ =Earned premium during the previous 12 month for non-life insurance obligations, without deducting the premium ceded to reinsurance
- Exp_{ul} =Amount of annual expenses incurred during the previous 12 months in respect life insurance where the investment risk is borne by the policyholders.
- $BSCR$ = Basic SCR

All the aforementioned inputs should be available for the last economic period and the previous one, in order to compute their last annual variations.

The capital charge as described in the CEIOPS QIS[6] 5 document:

$$SCR_{op} = \min\{0.30 \times BSCR; Op\} + 0.25 \times Exp_{ul}$$

where Op =Basic operational risk charge for all business other than life insurance where the investment risk is borne by the policyholders

$$Op = \max\{Op_{premiums}; Op_{provisions}\}$$

[6] Quantitative Impact Study.

where $Op_{premiums} = 0.04 \times \left(Earn_{life} - Earn_{life-ul}\right) + 0.03 \times Earn_{nl}$

$$+\max\{0; 0.04 \times (Earn_{life} - 1.1 \times pEarn_{life} - \left(Earn_{life-ul} - 1.1 \times pEarn_{life-ul}\right))\}$$
$$+ \max\{0; 0.03 \times (Earn_{nl} - 1.1 \times pEarn_{nl}\}$$

and $Op_{provisions} = 0.0045 \times max\{0; TP_{life} - TP_{life-ul}\} + 0.03 \times max\{0; TP_{nl}\}$

The drawback of the Standard Formula despite being easy to compute and to apply is that the Operational Risk SCR does clearly not reflect the risk profile of the company as it is a factor based approach.

In the EIOPA Report on the fifth Quantitative Impact Study for Solvency II (march 2011), the companies that have developed an internal model have around 5% of their SCR allocated in Operational Risks.

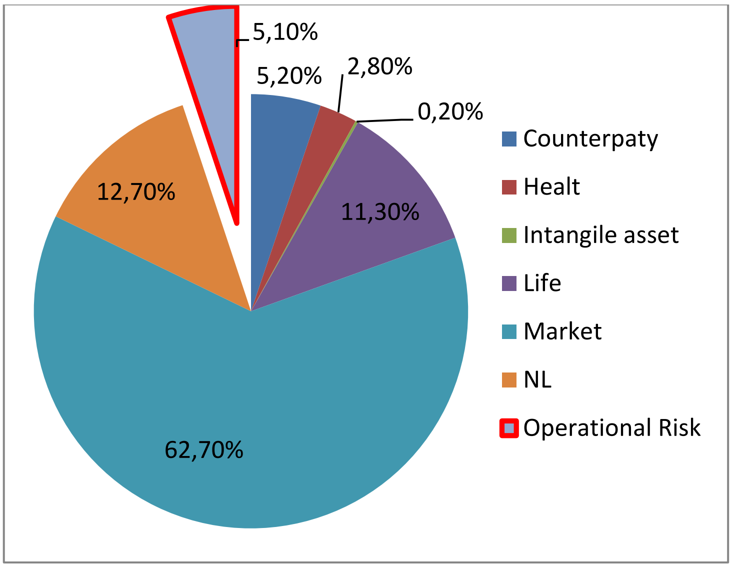


Figure 1: Repartition of the SCR among the main risks

2. Internal Model

In the Solvency II guidelines, the choice of using the standard formula or not is left to the Insurer. The alternative to the standard formula is the internal model. But making the choice is not straightforward. The implementation of an internal model can be very time-consuming and costly. Many insurance companies refer to the banking industry to build the internal model in operational risk.

In Basel II, a sophisticated and complex approach named the Advanced Measurement Approach (AMA) is introduced. It is the only one risk-sensitive approach allowed for operational risk. One of the main advantages of this approach is to improve the risk management processes thanks to a deeper risk analysis. In addition, the AMA can allow a reduction of the operational risk capital charge. Three approaches will be presented in this thesis. The Bayesian Approach and the Loss Distribution Approach will be described in this section. The Scenario Based Approach (SBA), which is AXA's internal model approach, will be more detailed in the next chapter.

a) Bayesian Approach

The Bayesian inference is well-suited statistical tool in using at the same time expert opinions and historical data:

- Expert judgments are incorporated into the internal model via specifying prior distributions; the expert specifies the distribution for the model parameter.
- The probabilistic theory used is derived from Bayes' theorem:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

The concept is to create a Bayesian network, which is a probabilistic causal graph, derived from the above theorem. Each node in the graph represents variables of interest, while the arrows between the nodes are the causal link between the corresponding random variables. The experts define with this process the Bayesian network. The parameters are then determined either with statistical data or still with expert judgments. Grouping all the created Bayesian networks allows us to create a global network: the aggregate loss is simply the sum of all losses.

One of the advantages of this approach is the link of causality between the variables. We are able to avoid a diversification issues: there is no need to estimate correlations between the risks. By construction, the network computes the aggregated loss by going through every random variable (that is defined with an estimated conditional distribution).

A drawback of the Bayesian approach is that constructing one is time-consuming and technical.

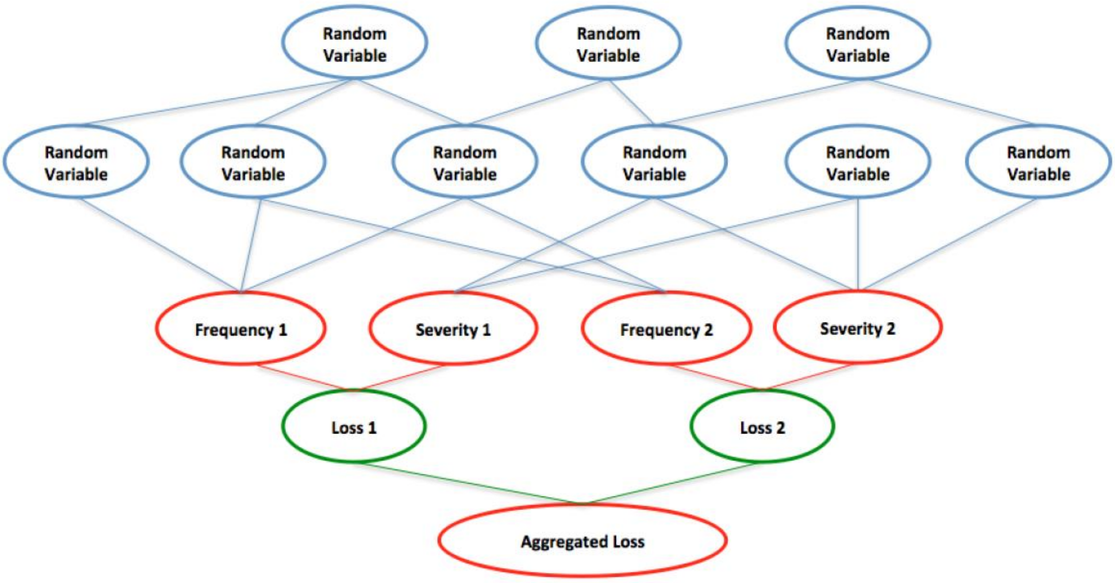


Figure 2 Bayesian Network Graph

b) Loss Distribution Approach (LDA)

This approach would be by far the most used and popular. The principle is fitting statistical distributions with available historic data (either internal or external). More precisely, we compute a frequency and severity distribution with which we recreate a total loss distribution.

$$S = \sum_{i=1}^{N} X_i$$

Where N is the random variable representing the frequency $\forall i \in [\![1, N]\!]$ $X_i \sim X$ where X is the random variable representing the severity and S is the total loss. This approach will be more detailed in the last chapter of this thesis.

c) Scenario based approach

This approach principle is based on scenario analysis: it consists in identifying all the operational risks that could exist in the company, and quantifying their frequency and severity distribution. The frequency module is divided in various frequency criteria meanwhile the financial impacts are divided into several impact types. The experts assess each of the components of the scenarios so that an estimation of the average frequency as well as an estimation of the average and extreme impacts are given. This method heavily depends on the expert judgment, but internal and external data can be used in the risk assessment. A quantile-matching approach is then used to calibrate the frequency and severity parameters. The aggregate loss is computed with a Monte Carlo approach, with a diversification effect taken into account. It is a forward-looking approach and it better represents the risk profile of a company. The SBA will be more detailed in the next chapter.

3. Operational Risk Data

Operational Risk modeling is a challenging task due to the lack of data and its heterogeneity. Even with more than one approach implementing an internal model is not trivial. The use of all the available information is needed.

a) Four data element

In Basel Committee's Supervisory Guidelines for the Advanced Measurement Approach (AMA) paper, it is said that four data elements are required for the quantification of the Operational Risk, which are:

- Internal Loss Data – ILD:

The ILD is expected to be used in the operational risk measurement system to assist in the estimation of loss frequencies, to inform the severity distribution(s) to the extent possible and to serve as an input into scenario analysis.

- External Data – ED:

ED are valuable to estimate the tail of an operational risk distribution. They are also essential input in scenario analysis and they provide valuable information about the losses experienced by other Insurance company.

- Scenario Analysis – SA:

Scenario Analysis focuses on the examination events that can be qualified of "low frequency – high impact". Experts are involved in the risk assessment process, it encourages managers in the evaluation of operational risks; it is a key step for the link between the measurement and the management of operational risk.

- Business Environment and Internal Control Factors – BEICFs:

"BEICFs are operational risk management indicators that provide forward-looking assessments of business risk factors as well as a bank's internal control environment. […].

BEICFs are commonly used as an indirect input into the quantification framework and as an ex-post adjustment to model output. Ex-post adjustments serve as an important link between the risk management and risk measurement processes […].”

The four sources of information are essential in the conception of an internal model. They are also a needed step in the validation of the internal model by the regulator.

b) Focus on expert judgment

With the difficulty of having coherent data, the expert judgment is often used as an alternative. A clear example is the Bayesian approach, where expert judgments are combined with statistical tools to calibrate severity distribution.

Expert judgments are useful in the quantification assessment, but it has to be reminded that biased expert judgments cannot be avoided. We can recognize different origin of biased expert judgments:

- Overconfidence:

Experts could overestimate the precision of their knowledge and over-rely on limited evidence.

- Optimism and wishful thinking:

There is a danger of experts having unrealistically bright forecast. In particular, managers tend to be too optimistic about the effectiveness of their action plans related to extremely rare and adverse event.

- Sample size neglect:

Samples can be considered appropriate without taking into consideration their size. Experts could estimate parameters from a sample without taking into account its size.

- Group polarization:

This refers to the tendency of groups to adopt more extreme positions that dominant individual members propose.

- Motivation bias:

Internal conflicts where participants have a conflict of interest related to the results of a scenario analysis workshop. Either for personal or for result oriented purpose.

- Availability or memory:

People often tend to overestimate the likelihood of recent incidents. Excessive attention can be attributed to extreme events and frequent events tend to be neglected.

- Framing:

Inconsistent choices or predictions for the same problem when phrased differently.

- Representativeness:

Expert can misinterpret the links between two events and that often lead to insufficient attention to individual probabilities of occurrence.

- Structural bias:

The expert is fully influenced by the scenario pre-existing design and does not bring sophistication from his own knowledge.

Chapter 3: Scenario Based Approach (SBA)

Among the three main operational risk internal model approaches, two were presented earlier in this thesis. The approach used for AXA's internal model is the Scenario Based Approach (SBA). It is a forward-looking approach and reflects the risk profile with adequacy. AXA's internal model takes into account Group and local specificities and addresses the Solvency II standard formula shortcoming. This approach relies on expert judgments, which are all back-tested. The scenarios allow testing the capacity of the company to face extreme event or natural disaster: some events taken into account in the scenarios have yet to occur. The use of the four data elements of the operational risk framework by experts is common. Plus, any complementary elements such as key risk indicators contribute greatly to ensure and enhance the accuracy of the experts and their ability to assess risks.

Let's do not forget the main purpose of an internal model: reflect the company's risk profile in the SCR. Plus, this is particularly true for Operational risks, as the Standard formula is not taking into account any factor based on an operational risk criteria.

1. Risk measures

Before starting to compute any capital charge, a Risk Measure has to be chosen.

Let's denote$\rho: \mathcal{L} \to \mathbb{R} \cup \{+\infty\}$, $\mathcal{L} = \{Set\ of\ Risks\}$, by definition a function that returns a real number that quantifies the risk. To be a coherent risk measure function, some properties must be met by the function:

- *(TI)* Translation Invariance $\rho(X + a) = p(X) + a, \;\; \forall X \in \mathcal{L}\ and\ a \geq 0$
- *(PH)* Positive homogeneity $p(\lambda X) = \lambda p(X), \;\; \forall X \in \mathcal{L}\ and\ \lambda \geq 0$
- *(M)* Monotonicity $X \leq Y \;\; \to \;\; \rho(X) \leq \rho(Y), \;\; \forall X, Y \in \mathcal{L}$
- *(SA)* Subadditivity $\rho(X + Y) \leq \rho(X) + \rho(Y), \;\; \forall X, Y \in \mathcal{L}$

The most commonly used risk measures are the Value-at-Risk ($VaR_{99.5\%}$) and the Tail Value-at-Risk ($TVaR_{99.5\%}$).

In the Solvency II Directives, the VaR_{α} is recommended to compute the economic capital charge.

<u>Definition of the VaR_{α}:</u>

$$VaR_{\alpha}(X) = \inf\{x \in \mathbb{R}, \mathbb{P}(X \leq x) \geq \alpha\},$$
$$where\ \mathbb{P}(X \leq x)\ is\ the\ cumulative\ distribution\ function\ of\ the\ risk.$$

The VaR_α is not a coherent risk measure because some of the above properties are not met. In fact, the VaR_α does not have the (SA) property. There are many discussions on whether to use or not the VaR_α as a risk measure. Plus, we have a threshold that indicates us the level of capital charge, but we don't have any information on how the distribution behaves above this threshold. Nevertheless, we have a coherent risk measure with the $TVaR_\alpha$.

By definition: $TVaR_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 VaR_\xi(X)d\xi$, it represents the average loss above the VaR_α. Nevertheless, the $TVaR_\alpha$ is difficult to compute and to interpret.

As required by the Solvency II Directives, AXA uses the VaR_α as the risk measure to compute its Operational risk Economic capital or Operational risk SCR. It is based on a one year horizon, with a 99.5% level.

The computation of the Operational risk figure is based on a Monte Carlo approach and we have that:

$$SCR = VaR_{99.5\%}$$

where $VaR_{99.5\%}$ is the $quantile(99.5\%)$ of the simulated aggregated loss.

2. Methodology

A single methodology is adopted by all the entities of the Group. In Solvency II, and more precisely in the Pillar II, the importance of insurance companies to identify their risk is highlighted. The identification of the risks is an important and non-negligible step of the quantification of the Operational risk SCR. AXA's methodology is in the spirit of Solvency II Directives: it encourages the entities to build their own risk profile.

The SBA is complemented by additional elements such as the four data elements (Internal losses, External losses, and Business Environment and Internal control factors).

The main objectives of AXA's methodology are to:

- Identify the most significant Operational risks, detect the sources and collect the losses
- Quantify the frequency and severity associated to each main Operational risk in compliance with the Solvency II regulatory requirements
- Mitigate by creating concrete action plans that can be activated to face catastrophic events
- Report and monitor to ensure the concordance between the risk profile and the risk appetite framework.

a) Scenario calibration

Operational risk modeling is based on a frequency / severity approach. The annual loss is equal to the sum of random number of losses.

Frequency measurement

The frequency describes the number of events that occur given a period of time. In Solvency II, it is a one-year period that is preconized. A risk's frequency is measured considering three different elements as contributing to a potential loss such as:

- a triggering factor: the cause or sources that make the risk occur
- a resource failure: it represents the vulnerability of the company to a specific type of failure, hence given the failure considerer as triggering the risk
- a loss generating failure: events producing a financial loss

The combination of all these elements will give the frequency of the scenario.

Severity measurement

The severity assessment is performed in order to give the amount of an incurred loss. Given the scarcity of data and the diversity of the risks, a global approach that can be replicated throughout the Group is needed. A straightforward approach is to consider parametric families of distributions, and calibrate them given assessed fitting points. Moments, mode and quantiles could be considered. The approach for the self-risk assessment provides a good illustration of the advantages and difficulties brought by the structure of parameterization. Let's first describe it so we may after bring light into aspects that need our attention.
The diversity of the risks considered has to be controlled by a limited number of parameters giving the shape of the loss distribution. When conceiving the internal model, it was decided that three impact levels would be given to assess the severity distribution. With most of distribution laws used in industry being determined by two parameters, the choice of giving an extra level was motivated by the fact that expert opinion would need more than one value to measure risk in the tails.

The three levels of impacts used are:

- Typical impact: it is the mode of the loss distribution that is to say the most frequent value the impact will take.
- Serious impact: it is the $quantile(\beta)$ of the severity distribution, i.e. the worst loss out of $\frac{1}{1-\beta}$ losses.

- Extreme impact: it is the $quantile(\gamma)$ of the severity distribution, i.e. the worst loss out of $\frac{1}{1-\gamma}$ losses.

These three levels of impact are the one used in AXA's methodology to calibrate the severity distribution. The next chapter will describe the calibration of the severity distributions.

b) Quantification

The risk measurement is important in order to rank with more precision the risk identified during the identification phase and to provide rationale for capital allocation. It allows the upper management to consider all the time the risk of having extreme events sometimes not catch in other approach.

Quantification of a scenario

The most important and critical risks of each entity are identified during the identification process. The common practice is to compute the annual loss distribution with a Monte Carlo approach. The frequency and the severity are modeled independently.

We have what we call a "*modèle collectif*" when we combine the frequency and the severity distribution calibrated. We have the following formula:

$$S = \sum_{i=1}^{N} X_i$$

where N corresponds to the number of events that occur during the considered scope of time (in our case, it is a one-year period)

X_i corresponds to the amount of each loss event. We have that all the individual losses are independent and identically distributed (iid) according to X

In this model, the frequency and severity are assumed independent. It allows us to write the mean and variance of the annual losses:

$$\mathbb{E}(S) = \mathbb{E}(X)\mathbb{E}(N)$$

$$\mathbb{V}(S) = \mathbb{E}(N)\mathbb{V}(X) + \mathbb{V}(N)\big(\mathbb{E}(X)\big)^2$$

We can deduce the cumulative distribution function of the annual loss:

$$F_S(s) = \begin{cases} \displaystyle\sum_{n=1}^{+\infty} \mathbb{P}(N = n)F_X^{\otimes n}(x), & when\ x > 0 \\ \mathbb{P}(N = 0), & when\ x = 0 \end{cases}$$

where $F_X^{\otimes n}$ is the cumulative distribution function of n losses $(X_1, X_2, \dots, X_n)$ obtained by the convolution of order n if the severity distribution[7].

It is common to use a Poisson distribution for the random variable N. Depending on how the scenario is built, other counting distribution can be used such as the Bernoulli or Binomial distribution. Also, depending on the characteristics of the assessed impacts, the severity distribution is assumed to follow a lognormal or Pareto distribution. Again, other distributions such as the Weibull or exponential can be used.

These risks are aggregated with Monte Carlo methods. We estimate the capital charge coming from operational risk. We have a diversification effect due to the correlation coefficients that measure the dependence of variables.

The quantification of operational risk with a SBA will be detailed in the next chapter.

[7] We remind that for X, Y two independent and continuous random variables, with f and g their respective density function, then $Z = X + Y$ has for density function $h(x) = \int_{-\infty}^{+\infty} g(y) f(x-y)\, dy = f * g$.

Chapter 4: Quantification of scenarios

The first step in quantifying operational risk is the scenario modeling. We use as explained in the previous chapter, a frequency-severity approach. The experts assess the three impacts (typical, serious and extreme), and we fit a distribution, by calibrating it with the three impacts. This chapter presents a calibration based on a quantile-matching method. A study on the frequency threshold is included in this chapter to insure a positive $VaR_{99.5\%}$. In fact, in some cases, we had a null $VaR_{99.5\%}$. The approach used in AXA for computing the total loss distribution is also presented in this chapter.

1. Frequency distributions:

We define a Loss Frequency Distribution and a Loss Severity Distribution for each operational risk. The parameters of the distributions are estimated from the risk assessments results. We combine the two distributions with Monte Carlo methods to obtain the estimated Loss distribution of the operational risk.

Each operational risk annual loss can be defined by:

$$S = \sum_{i=1}^{N} X_i$$

where N corresponds to the number of events that occur during the considered scope of time (in our case, it is a one-year period)

X_i corresponds to the amount or the magnitude of each loss event. We have that all the individual losses are independent and identically distributed (iid) according to X

a) Poisson

The Poisson distribution is used to model the number of events that occurs in a given scope of time (in our case, it is a one-year period. The parameter $\lambda \in \mathbb{R}_+^*$ of a Poisson distribution corresponds to the mean and the variance of the distribution. In the insurance industry, the Poisson distribution is often used when the mean and the variance of the frequency seen in the data are close.

$$\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, with\ k \geq 0$$

$$\mathbb{E}(X) = \lambda\ and\ \mathbb{V}(X) = \lambda$$

b) Bernoulli

It is a frequency distribution used to model a risk that cannot happen more than once in a year. It can be for example a fine by the regulator on an erroneous tax report. It has one parameter $p \in [0,1]$, the probability of occurrence of the risk. We can represent the Bernoulli distribution by a coin toss.

$$\mathbb{P}(X = k) = \begin{cases} p & if\ k = 1 \\ 1 - p & if\ k = 0 \\ 0 & else \end{cases}$$

$$\mathbb{E}(X) = p\ and\ \mathbb{V}(X) = p(1 - p)$$

c) Binomial

It has two parameters: $(n, p) \in \{\mathbb{N}_+^*, [0,1]\}$. It is an appropriate choice to model extreme events where there is relatively little variation and where the risk will not occur more than n times. In addition, a Binomial distribution constitutes an adequate choice if the frequency variance is smaller than the mean. It a generalized[8] Bernoulli distribution.

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, for\ k = 0,1,2, \dots, n$$

$$\mathbb{E}(X) = np\ and\ \mathbb{V}(X) = np(1 - p)$$

d) Negative Binomial

Like the binomial distribution, the Negative Binomial has two parameters $(r, p) \in \{\mathbb{R}_+^*, [0,1]\}$, thus making it more flexible than the Poisson distribution. The Negative Binomial distribution is an appropriate choice if the frequency variance is larger than the mean (often seen in empirical data). It is used to model the frequencies subject to a high degree of variability.

[8] A binomial distribution can be defined as the sum of n independent Bernoulli distribution of parameter p.

$$\mathbb{P}(X=k)=\binom{k+r-1}{k}p^k(1-p)^r, for\ k=0,1,2,\dots,n$$

$$\mathbb{E}(X)=\frac{pr}{1-p}\ and\ \mathbb{V}(X)=\frac{pr}{(1-p)^2}$$

e) Fixed Frequency

Usually, the frequency distribution used in operational risk modeling is one of the above distributions. We can also mention the "Fixed Frequency" distribution. With the fixed frequency, we have that the mean is equal to a constant with a null variance. It is an appropriate choice if we know the frequency of the event. We can note that when the fixed frequency is setup at 1, the annual loss is only determined by the severity distribution.

2. Severity distributions:

The estimation of the severity distribution consists in giving the amount of an incurred loss. Many approaches can be adopted to estimate a severity distribution as the Maximum of Likelihood Estimator (MLE) or with a method of moments approach. In this thesis, only a quantile-approach is presented as it is the AXA's approach to calibrate the severity distributions.

First, let's consider the three levels of impact:

- Typical impact: it is the mode of the loss distribution that is to say the most frequent value the impact will take.
- Serious impact: it is the $quantile(\beta)$ of the severity distribution, i.e. the worst loss out of $\frac{1}{1-\beta}$ losses.
- Extreme impact: it is the $quantile(\gamma)$ of the severity distribution, i.e. the worst loss out of $\frac{1}{1-\gamma}$ losses.

In this section, we will calibrate the severity distribution parameters with these three assessed impacts. As the method of calibration here is a quantile-matching, we will retrieve the analytical formulas of the quantile of severity distributions. All the distributions that can be used in Operational Risk modeling won't be presented here. Only the most used will be. Some other distribution such as the exponential or gamma distribution can be used as severity distribution.

a) Lognormal

The lognormal distribution is a right heavy-tailed distribution, and is particularly appropriate the low frequency / high severity events.

Let's denote X a lognormal distribution. We have that

$$\ln(X) \sim \mathcal{N}(\mu, \sigma^2)\,, with\ \mu\ the\ mean\ and\ \sigma\ the\ standard\ deviation, (\mu, \sigma) \in (\mathbb{R}, \mathbb{R}_+^*)$$

$$\Leftrightarrow \frac{\ln(X) - \mu}{\sigma} \sim \mathcal{N}(0,1)$$

Since we have quantile tables for the $\mathcal{N}(0,1)$ distribution, we retrieve analytical formulas of the lognormal distribution X:

$$\mathbb{P}\left(\frac{\ln(X) - \mu}{\sigma} \leq q_\alpha^{\mathcal{N}(0,1)}\right) = \alpha \Leftrightarrow \mathbb{P}\left(X \leq e^{\sigma q_\alpha^{\mathcal{N}(0,1)} + \mu}\right) = \alpha$$

$$thus, quantile(X, \alpha) = e^{\sigma q_\alpha^{\mathcal{N}(0,1)} + \mu}$$

With the three impacts taken in account for the calibration, we have:

$$\begin{cases} serious = e^{\sigma q_\beta^{\mathcal{N}(0,1)} + \mu} \\ extreme = e^{\sigma q_\gamma^{\mathcal{N}(0,1)} + \mu} \end{cases}$$

By definition, the density of a lognormal distribution is:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\ln(x) - \mu)^2}$$

The typical impact is the mode of the severity distribution, i.e. the loss that will occur the most of the time. It is also the maximum of the density function.

$$mode = e^{\mu - \sigma^2}$$

We have then the three impacts estimated in function of the severity distribution parameters:

$$\begin{cases} mode = e^{\mu - \sigma^2} \\ serious = e^{\sigma q_\beta^{\mathcal{N}(0,1)} + \mu} \\ extreme = e^{\sigma q_\gamma^{\mathcal{N}(0,1)} + \mu} \end{cases} \Leftrightarrow \begin{cases} mode = e^{\mu - \sigma^2} \\ \dfrac{\ln(serious) - \mu}{\sigma} = q_\beta^{\mathcal{N}(0,1)} \\ \dfrac{\ln(extreme) - \mu}{\sigma} = q_\gamma^{\mathcal{N}(0,1)} \end{cases}$$

Let's solve the system of three equations and two unknowns. We could have now the parameters of the lognormal distribution. But the calibration of the parameters will be based on three impacts, and not two in order to better use the expert's judgments. We introduce a weighting factor $a \in [0,1]$ that is the weight allocated to the serious impact:

$$a \times \left(\frac{\ln(serious) - \mu}{\sigma}\right) + (1-a) \times \left(\frac{\ln(extreme) - \mu}{\sigma}\right) = a \times q_\beta^{\mathcal{N}(0,1)} + (1-a) \times q_\gamma^{\mathcal{N}(0,1)}$$

$$thus, \mu = a \times \ln(serious) + (1-a) \times \ln(extreme) - \sigma\left(a \times q_\beta^{\mathcal{N}(0,1)} + (1-a) \times q_\gamma^{\mathcal{N}(0,1)}\right)$$

$$We\ have, \begin{cases} \mu = \ln(mode) + \sigma^2 \\ \mu = a \times \ln(serious) + (1-a) \times \ln(extreme) - \sigma\left(a \times q_\beta^{\mathcal{N}(0,1)} + (1-a) \times q_\gamma^{\mathcal{N}(0,1)}\right) \end{cases}$$

The two equations combined give the following equation:

$$\sigma^2 + \sigma\left(a \times q_\beta^{\mathcal{N}(0,1)} + (1-a) \times q_\gamma^{\mathcal{N}(0,1)}\right) + \ln\left(\frac{mode}{extreme}\right) + a \times \ln\left(\frac{extreme}{serious}\right) = 0$$

The second equation is a 2nd degree polynomial of σ. The roots are:

$$\sigma_\pm = \frac{-\left(a \times q_\beta^{\mathcal{N}(0,1)} + (1-a) \times q_\gamma^{\mathcal{N}(0,1)}\right) \pm \sqrt{\Delta}}{2}$$

$$where\ \Delta = \left(a \times q_\beta^{\mathcal{N}(0,1)} + (1-a) \times q_\gamma^{\mathcal{N}(0,1)}\right)^2 - 4\left(\ln\left(\frac{mode}{extreme}\right) + a \times \ln\left(\frac{extreme}{serious}\right)\right)$$

Finally we have the parameters of the lognormal distribution that are defined as follow:

$$\begin{cases} \mu = \ln(mode) + \sigma^2 \\ \sigma = \dfrac{-\left(aq_\beta^{\mathcal{N}(0,1)} + (1-a)q_\gamma^{\mathcal{N}(0,1)}\right) + \sqrt{\left(aq_\beta^{\mathcal{N}(0,1)} + (1-a)q_\gamma^{\mathcal{N}(0,1)}\right)^2 - 4\left(\ln\left(\frac{mode}{extreme}\right) + a\ln\left(\frac{extreme}{serious}\right)\right)}}{2} \end{cases}$$

b) Pareto

The Pareto distribution is a very popular distribution to model low frequency / high severity events. It is a fat right-tailed distribution (also called heavy tailed distribution). Plus, the existence of a threshold allows us to better capture extreme events that are over that threshold.

Let's denote X a Pareto distribution. We have the density function that is defined as following:

$$f(x) = k\frac{x_{min}^k}{x^{k+1}}\mathbb{I}_{\{x>x_{min}\}}, \qquad with\ (x_{min}, k) \in (\mathbb{R}, \mathbb{R}_+^*)$$

Let's derive the density function:

$$f'^{(x)} = kx_{min}^k \frac{-(k+1)}{x^{k+2}}\mathbb{I}_{\{x>x_{min}\}}$$

As both parameters x_{min} and k are non-negative (>0) and that with $x \in [x_{min}, +\infty[$ we have $f'(x) < 0, mode = x_{min}$.

Let's apply the quantile-matching approach with the Pareto distribution.

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} k\frac{x_{min}^k}{y^{k+1}}\mathbb{I}_{\{y>x_{min}\}}\,dy = \int_{x_{min}}^{x} k\frac{x_{min}^k}{y^{k+1}}\,dy = 1 - \left(\frac{x_{min}}{x}\right)^k$$

Thus, if $q_\alpha^{\mathcal{P}}$ the quantile of level α, then

$$F(q_\alpha^{\mathcal{P}}) = \alpha \Leftrightarrow 1 - \left(\frac{x_{min}}{q_\alpha^{\mathcal{P}}}\right)^k = \alpha \Leftrightarrow \ln(1-\alpha) = k\ln\left(\frac{x_{min}}{q_\alpha^{\mathcal{P}}}\right) \Leftrightarrow k = \frac{\ln(1-\alpha)}{\ln\left(\frac{x_{min}}{q_\alpha^{\mathcal{P}}}\right)}$$

We apply the same methodology with the introduction of a weighting factor between the serious and extreme events.

$$\begin{cases} k\ln\left(\frac{x_{min}}{q_\beta^{\mathcal{P}}}\right) = \ln(1-\beta) \\ k\ln\left(\frac{x_{min}}{q_\gamma^{\mathcal{P}}}\right) = \ln(1-\gamma) \end{cases} \Rightarrow k = \frac{a\ln(1-\beta) + (1-a)\ln(1-\gamma)}{a\ln\left(\frac{x_{min}}{q_\beta^{\mathcal{P}}}\right) + (1-a)\ln\left(\frac{x_{min}}{q_\gamma^{\mathcal{P}}}\right)}$$

Finally we have the parameters of the Pareto distribution that are defined as follow:

$$\begin{cases} x_{min} = typical \\ k = \dfrac{a\ln(1-\beta) + (1-a)\ln(1-\gamma)}{a\ln\left(\frac{x_{min}}{q_\beta^{\mathcal{P}}}\right) + (1-a)\ln\left(\frac{x_{min}}{q_\gamma^{\mathcal{P}}}\right)} \end{cases}$$

c) Weibull

It is a distribution of the sub-exponential family, i.e. a distribution whose tail decays slower than the exponential distribution (in fact, we can retrieve an exponential distribution with a special set of parameters of a Weibull distribution).

Let's denote X a Weibull distribution. We have the density function that is defined as following:

$$f(x) = \begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k} \mathbb{I}_{\{x\geq 0\}}, k > 1 \\ \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k} \mathbb{I}_{\{x>0\}}, k < 1 \end{cases}, with\ (\lambda, k) \in (\mathbb{R}_+^*, \mathbb{R}_+^*)$$

We only consider the case where $k < 1$. In this case, the mode of the distribution is not defined. We use then only the serious and extreme impacts assessed by the expert to calibrate the model.

The cumulative distribution function is defined as follow:

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} \frac{k}{\lambda}\left(\frac{y}{\lambda}\right)^{k-1} e^{-\left(\frac{y}{\lambda}\right)^k} \mathbb{I}_{\{y>0\}}\, dy = \int_{0}^{x} \frac{k}{\lambda}\left(\frac{y}{\lambda}\right)^{k-1} e^{-\left(\frac{y}{\lambda}\right)^k} dy = 1 - e^{-\left(\frac{x}{\lambda}\right)^k}$$

We see that when $k = 1$, we have an exponential distribution of parameter $\lambda \in \mathbb{R}_+^*$.

Thus, if q_α^W the quantile of level α, then

$$F(q_\alpha^W) = \alpha \Leftrightarrow 1 - e^{-\left(\frac{q_\alpha^W}{\lambda}\right)^k} = \alpha \Leftrightarrow \ln(1-\alpha) = -\left(\frac{q_\alpha^W}{\lambda}\right)^k \Leftrightarrow q_\alpha^W = \lambda(-\ln(1-\alpha))^{\frac{1}{k}}$$

We have a system of equation:

$$\begin{cases} q_\beta^W = \lambda(-\ln(1-\beta))^{\frac{1}{k}} \\ q_\gamma^W = \lambda(-\ln(1-\gamma))^{\frac{1}{k}} \end{cases} \Leftrightarrow \begin{cases} \ln(serious) = \ln(\lambda) + \frac{1}{k}(-\ln(1-\beta)) \\ \ln(extreme) = \ln(\lambda) + \frac{1}{k}(-\ln(1-\gamma)) \end{cases}$$

$$\Rightarrow \begin{cases} \lambda = \dfrac{serious}{(-\ln(1-\beta))^{\frac{1}{k}}} \\ \lambda = \dfrac{extreme}{(-\ln(1-\gamma))^{\frac{1}{k}}} \end{cases}$$

We can deduce the following calibration for the parameters of the Weibull distribution:

$$\begin{cases} \lambda = \dfrac{serious}{(-\ln(1-\beta))^{\frac{1}{k}}} \\ k = \dfrac{\ln(-\ln(1-\beta)) - \ln(-\ln(1-\gamma))}{\ln(serious) - \ln(extreme)} \end{cases}$$

We can with this approach take the weight of the extreme event into account in the calibration of the parameter k.

3. Goodness of fit

The severity distributions used in operational risk modeling usually have two parameters. It is the case for the lognormal and Pareto distributions. But how do we verify the robustness of the calibrated severity distribution? Since discrepancies are inherent in the calibration of the model, we will compute the relative error between the assessed impacts by the experts and the impacts calibrated by the model. Also, before computing any predicted values, we have to fix a value for the weighting factor introduced to take into account both the serious and extreme impacts in the calibration of the severity distributions (lognormal and Pareto generally).

The relative error[9] here is our measure:

$$relative\ error = \frac{absolute\ error}{actual\ value} = \frac{|predicted\ value - actual\ value|}{actual\ value}$$

$$where \begin{cases} predicted\ value = impact\ calibrated\ by\ the\ model \\ actual\ value = impact\ assessed\ by\ the\ expert > 0 \end{cases}$$

To measure the accuracy of the model, we use the SSRE: Sum of Squared Relative Error

$$SSRE = \sqrt{\sum_{i=1}^{n}(relative\ error)^2} = \sqrt{\sum_{i=1}^{n}\left(\frac{|predicted\ value - actual\ value|}{actual\ value}\right)^2}$$

Even if we have assessed three impacts by the expert, we have that $n = 2$. In fact, we compute the relative error only with the serious and extreme impacts. We deduce the predicted values with the parameters of the severity distribution that has been calibrated with the assessed impacts.

Let's take a Pareto distribution of parameters $(x_{min}, k) \in (\mathbb{R}, \mathbb{R}_+^*)$ as an example. The actual values are the following:

- $q_\beta^{\mathcal{P}} = exp\left(\ln(x_{min}) - \frac{1}{k}\ln(1-\beta)\right)$
- $q_\gamma^{\mathcal{P}} = exp\left(\ln(x_{min}) - \frac{1}{k}\ln(1-\gamma)\right)$

So the SSRE is computed as following:

[9] The relative error is used here because we compare approximations of numbers of widely differing size.

$$SSRE_{\mathcal{P}} = \sqrt{\left(\frac{\left|exp\left(\ln(x_{min}) - \frac{1}{k}\ln(1-\beta)\right) - serious\right|}{serious}\right)^2 + \left(\frac{\left|exp\left(\ln(x_{min}) - \frac{1}{k}\ln(1-\gamma)\right) - extreme\right|}{extreme}\right)^2}$$

We can use the SSRE to compute an error measure and compare the adequacy of both the lognormal and Pareto distribution. The distribution with the least SSRE would then be the "optimal distribution". Other criteria such as a finite mean and variance are added to determine the optimal distribution, especially for the Pareto distribution:

- the mean does not exist (i.e. is infinite) when the shape parameter $k \leq 1$
- the variance does not exist (i.e. is infinite) when the shape parameter $k \leq 2$

With the above criteria, we are able to determine an optimal distribution between a lognormal and a Pareto distribution.

4. Total loss distribution of a scenario

With a frequency distribution and severity distribution calibrated, we are able to compute the total loss distribution over one-year horizon. The Operational Risk SCR is the $quantile(99.5\%)$ of the total aggregated loss distribution. Many approaches were developed, but the aim of this master thesis is not to make a catalog, but to present AXA's approach: Monte Carlo simulations.

The most used approaches (there won't be presented as many theses have been presenting them already) for computing the total aggregated losses are:

- Panjer recursion algorithm
- Characteristic function inversion
- Extreme Value Theory or Extreme Value Analysis (EVA)

<u>Monte Carlo approach</u>

The Monte Carlo approach consists in simulating a high number of the total loss distribution when the frequency and severity distribution are already chosen.

- We simulate $n\ time$ the frequency distribution we have $F = (f_1, f_2, \dots, f_n)$
- For each element f_i of $F, i \in [\![1, n]\!]$, we simulate f_i losses that follow the severity distribution chosen and calibrated

$$\begin{cases} C_1(i) \\ \dots \\ C_{f_i}(i) \end{cases}$$

- We sum all these simulated losses to have 1 simulation of the total loss distribution:

$$S(i) = \sum_{k=1}^{f_i} C_k(i)$$

- We obtain n simulations of the total loss distribution
- The capital charge is then computed from these n total charge distribution: it is the Value-at-Risk at a 99.5% level ($VaR_{99.5\%}$).

Below is the R code to simulate a Poisson Lognormal scenario with a Monte Carlo approach:

```
MC=function(n, lambda, mu, sigma){
        a=rpois(n, lambda) #simulation of the frequency distribution
        X=sapply(a,function(a){sum(rlnorm[10](a,mu, sigma))}) #simulation of the severity distribution
        return (X)
}
```

The inputs are:

$$\begin{cases} n, the\ number\ of\ simulation\ in\ the\ Monte\ Carlo\ run \\ lambda, the\ frequency\ parameter \\ (mu, sigma)\ the\ severity\ parameters \end{cases}$$

The above R function allows us to simulate a vector of n total loss distribution simulations. We have to use the R function $quantile(.)$ to have the capital charge of the scenario. Note that in this approach, the parameters of the severity distribution must have already been calibrated according to the methodology presented earlier (see Chapter 5, section 1.b). An alternative code which calibrates the parameters of the severity distribution will be presented in the appendix.

5. Monte Carlo convergence

The Monte Carlo approach is useful in order to compute a total loss distribution. This section has for aim to demonstrate the convergence of Monte Carlo simulations. As the Monte Carlo approach is one of the most used to compute the total loss distribution, the stability and convergence of this approach is needed to ensure the robustness of the results< If we take $n = x$ simulations for each Monte Carlo run, we would like to show that this number of simulations ensure the convergence of the Monte Carlo run, i.e. the convergence of the Operational Risk SCR.

In order to examine the convergence of the $VaR_{99.5\%}$, we will compute the capital charge value using an increasing number of simulations and then plot the obtained figures: that would enable to graphically determine the number of Monte Carlo runs needed to have a

[10] If we want to simulate a Poisson-Pareto model, the packages “VGAM” must be installed.

satisfying convergence. Then, we will try to build an estimator for the interval of confidence by computing various times the capital charge value obtained with x simulations. A statistic descriptive table will let us focus on the variance of the sample.

First, we will plot the successive values of the capital charge during the simulation process. This analysis will keep the values of the capital charge computed based on a $1000 \times k$ simulation trip, with $k \in [\![1,1000]\!]$ (so that the last capital charge will be based on a simulation strip obtained with $N = 1\ 000\ 000$ simulations). We have reproduced the code computing these values.

```
X=SimulateScenario[11](lambda,mu,sigma,typef,types)
temp2=rep(0,step)
temp2[1]=STEC(X)
for (j in 2:step){
        Xtemp<-SimulateScenario(lambda,mu,sigma,typef,types)
        X=c(X,Xtemp)
        temp2[j]=STEC(X)
        rm(Xtemp)
}
```

If we take n = 500 000 simulations for a specific scenario with the following parameters:

- A Poisson distribution of parameter $\lambda = 0.04$
- A lognormal distribution of parameters $(\mu = 18.85, \sigma = 0.65)$

We obtain the following graph:

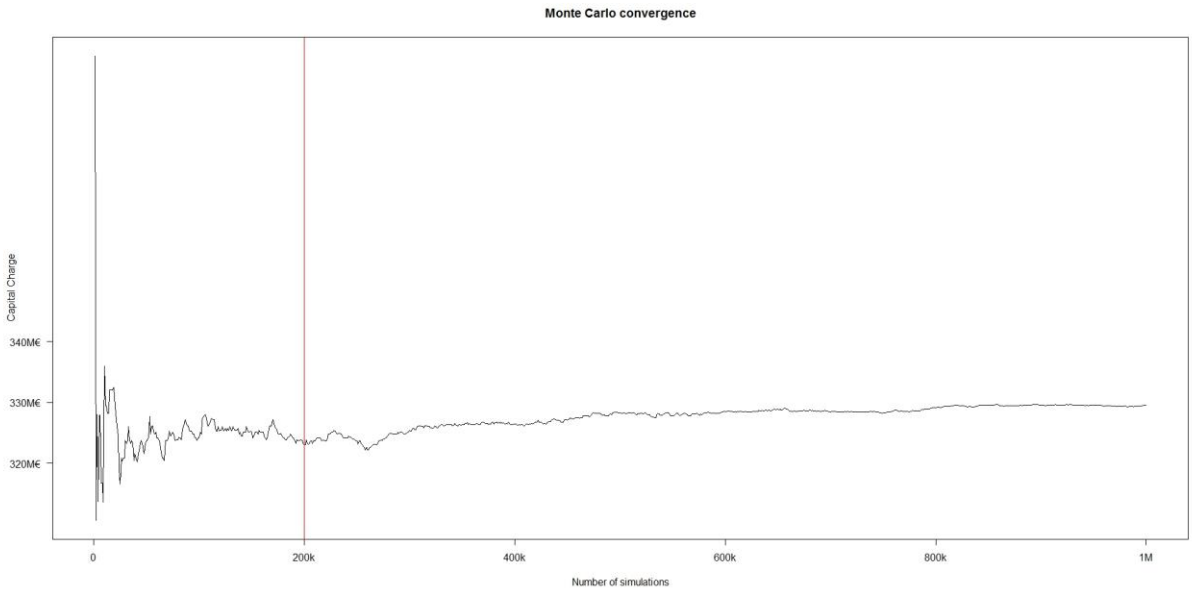


Figure 3 Graph of Monte Carlo convergence

We can see that the capital charge here reaches an acceptable level of convergence when the number of simulation is larger than 200'000. There, simulating standalone scenarios with 500'000 simulations ensures us to reach a stable value for the capital charge. We can see that the sample does not deviate too much from its mean.

---

[11] The full code is part of an internal tool of AXA developed during my internship. For confidentiality purpose it won't be presented here.

Minimum	1st Quartile	Median	3rd Quartile	Max	Mean	Std Dev
322 636 141	329 513 837	331 318 968	332 513 837	337 867 485	331 261 990	2 586 428

Table 3 Descriptive statistics 1'000 runs of Monte Carlo with 500k simulations

$$\frac{\sigma(capital\ charge)}{\mu(capital\ charge)} = \frac{2\,586\,428}{331\,261\,990} = 0.78\%$$

In order to see if 500'000 simulations is the adequate number, we simulated 1'000 times the STEC obtained with 100'000 to 750'000 simulations with a step of 50'000 each time. The results are the following:

100'000	150'000	200'000	250'000	300'000	350'000	400'000
1,76 %	1,51%	1,28%	1,15%	1,02%	0,95%	0,87%
450'000	**500'000**	**550'000**	**600'000**	**650'000**	**700'000**	**750'000**
0,82%	0,78%	0,77%	0,75%	0,73%	0,71%	0,69%

Table 4 Standard deviation of 1'000 runs of Monte Carlo with various simulations

As we can see, with the increase of the number of simulations, the sample is more accurate and the deviation from the mean is smaller. But from 500'000 to 750'000 simulations, the deviation to the mean does not decrease as fast as when there are between 200'000 and 500'000 simulations. In conclusion, we can define a threshold for the desired precision, which is the precision from the 200k simulations run. The choice of 500'000 simulations gives a good precision. The result is similar to the Value at Risk computed by the tool ORS (a difference of less than 0.50%), which allow us to conclude that 500k simulations is an adequate number of simulations to compute the STEC.

6. Frequency threshold

In the Solvency II directives, insurers are expected to prevent events that expected to occur once every two hundred years. We will set a threshold that ensures us a frequency level that allows us to compute a positive $VaR_{99.5\%}$.

A combination of Poisson-Lognormal model is generally considered as a standard market practice in terms of Operational Risk modeling. The parameters of the distributions used will have an impact on the capital charge required, i.e. the $VaR_{99.5\%}$. This study aims to define a threshold that ensures the positiveness of the $VaR_{99.5\%}$.

For some cases, the $VaR_{99.5\%}$ computed is null. Therefore, we would like to understand what drives this result by studying the following:

- The impact of the parameter of the frequency modeled with a Poisson distribution

- The impact of the seed
- The impact of the parameters of the severity modeled with a lognormal distribution
- The impact of the number of simulations in each Monte Carlo run

a) Frequency parameter

The first impact we study here is the frequency parameter. In fact, we want to determine if there is a link between λ, the frequency parameter and a null quantile of the distribution. Let's denote N a Poisson distribution of parameter λ. We want to demonstrate that

$$for\ \lambda < 1, \exists x \in]0;1[\ such\ that\ quantile(N,x) = 0$$

A Monte Carlo simulation will be zero when there is no event happening (i.e. when $N = 0$). Here, we are focusing on the risks for which N follows a Poisson distribution, that is: $N \sim \mathcal{P}(\lambda)$. We have:

$$\mathbb{P}(X = 0) = \mathbb{P}(N = 0) = e^{-\lambda}\frac{\lambda^0}{0!} = e^{-\lambda}$$

Let $(X_1, \dots, X_n)$ be the outcomes of a Monte Carlo run of n simulations. The $quantile(x)$ equal to zero is equivalent to the fact that there are $(\frac{1}{1-x} - 1)$ more simulations with an outcome equal to zero than non-zero outcomes. That is:

$$\{quantile(x, X_1, \dots, X_n) = 0\} \Leftrightarrow \sum_{k=1}^{n} \mathbb{I}\{X_k = 0\} > (\frac{1}{1-x} - 1)\sum_{k=1}^{n} \mathbb{I}\{X_k > 0\}$$

Yet:

$$\mathbb{P}(X = 0) > \left(\frac{1}{1-x} - 1\right)\mathbb{P}(X > 0)$$

$$\Leftrightarrow e^{-\lambda} > \left(\frac{1}{1-x} - 1\right)\sum_{k=1}^{+\infty} e^{-\lambda}\frac{\lambda^k}{k!}$$

$$\Leftrightarrow 1 > \left(\frac{1}{1-x} - 1\right)\sum_{k=1}^{+\infty}\frac{\lambda^k}{k!} \Leftrightarrow 1 > \left(\frac{1}{1-x} - 1\right)(\sum_{k=0}^{+\infty}\frac{\lambda^k}{k!} - \frac{\lambda^0}{0!})$$

$$\Leftrightarrow 1 + (\frac{1}{1-x} - 1) > (\frac{1}{1-x} - 1)\sum_{k=0}^{+\infty}\frac{\lambda^k}{k!}$$

$$\Leftrightarrow \frac{1}{1-x} > (\frac{1}{1-x} - 1)e^{\lambda}$$

$$\Leftrightarrow \lambda < \ln\left(\frac{\frac{1}{1-x}}{\frac{1}{1-x}-1}\right)$$

$$\Leftrightarrow \ \lambda < \ln\left(\frac{1}{x}\right)$$

$$\Leftrightarrow \lambda < -\ln(x)$$

Hence, we have a null $quantile(x)$ related to the frequency of the Poisson distribution. We have the following relation

$$\boxed{\lambda < -\ln(x) \ \Leftrightarrow \ quantile(x) = 0}$$

As an example, for the $VaR_{99.5\%}$, we have for $\lambda < -\ln(0.995) \approx 0.005012$ for which the value is always NULL. We have now a theoretical frequency threshold that is $\lambda = 0.005012$.

Besides, we could also derive what quantile will be null using a certain level of frequency. We have:

$$1 > \mathbb{P}(X = 0) > \left(\frac{1}{1-x} - 1\right)\mathbb{P}(X > 0)$$

$$\Leftrightarrow 1 > \frac{x}{1-x}\,\mathbb{P}(X > 0)$$

$$\Leftrightarrow \frac{1-x}{x} > \mathbb{P}(X > 0)$$

Then

$$\mathbb{P}(X = 0) > \frac{2x-1}{x}$$

If we take $\lambda = 0.00521$, we have $quantile(x, X_1, \dots, X_n) = 0$ for $x = \exp(-\lambda)$

So: $quantile(0.9948, X_1, \dots, X_n) = 0$ and $\mathbb{P}(X = 0) > \frac{2x-1}{x} = 0.9948$.

We can wonder with this value (99.48%) if the $VaR_{99.5\%}$ is affected or not. Let's denote n the number of simulations in a Monte Carlo run.

Due to the fact that we cannot have an infinite number of simulations in a Monte Carlo run, the $(99.48\% \times n) - th$ value of the re-ordered run is not 0 all the time. There are many parameters to take into account such as the seed used. Also, there is a high probability that $(99.48\% \times n)$ won't be an integer, which create variation in the theoretical null quantile defined previously.

We simulated a scenario with the following parameters (Poisson – Lognormal model):

$$\lambda = 0.0052\,, \qquad \mu = 17.1614\,, \ \ \sigma = 0.3693$$

We fixed all the parameters and we simulated 1000 times the $VaR_{99.5\%}$ (500k simulations in each run) with various frequency.

The first line is with $\lambda = 0.0052$. We have 96.10% of the 1000 runs for which the $VaR_{99.5\%}$ computed is superior to 0. It means that with $\lambda = 0.0052$ we can still have $VaR_{99.5\%} = 0$ even if we are above our theoretical threshold.

λ	x	VaR > 0
0.0052	99.481350%	96.10%
0.00521	99.480355%	97.20%
0.00522	99.479360%	97.90%
0.00523	99.478365%	98.00%
0.00524	99.477370%	98.80%
0.00525	99.476376%	98.90%
0.00526	99.475381%	99.20%
0.00527	99.474386%	99.60%
0.00528	99.473391%	99.40%
0.00529	99.472397%	99.80%
0.0053	99.471402%	99.80%
0.00531	99.470407%	99.70%
0.00532	99.469413%	100.00%
0.00533	99.468418%	99.90%
0.00534	99.467423%	100.00%
0.00535	99.466429%	99,90%
0.00536	99.465434%	100.00%
0.00537	99.464439%	100.00%
0.00538	99.463445%	100.00%
0.00539	99.462450%	100.00%
0.0054	99.4615%	100.00%
0.0055	99.4515%	100.00%
0.0056	99.4416%	100.00%
0.0057	99.4316%	100.00%
0.0058	99.4217%	100.00%
0.0059	99.4117%	100.00%
0.006	99.4018%	100.00%

Table 5 VaR simulations with various frequencies

We see that after $\lambda = 0.00536$, the $VaR_{99.5\%}$ computed for each 1000 runs is always superior to 0. We see here a difference between the theoretical threshold and the tests with various frequencies. We would like to know from where comes this variation of the threshold. For that, we will study the impact of the seed.

b) Impact of the seed

The choice of the seed might be important. According to the seed used, the results can vary a lot. In this study, the PRNG[12] used is the Mersenne Twister as the statistical software R uses it. In order to ensure that the $VaR_{99.5\%}$ will be always non negative with a frequency of $\lambda = 0.00536$, we tenfold the number of run, with now 10000 runs of Monte Carlo (with 500k simulations in each run). We run some tests on the value $\lambda = 0.00536$. We run 10000 times Monte Carlo run (500k simulations in each) of the scenario, and we studied the distribution of the value superior to 0.

Distribution of the percentage of null values value in each MC run for $\lambda = 0.00536$		
Minimum	**Max**	**Mean**
99,4258%	99,5036%	99,4656%

Table 6 Distribution of the null values in each MC run (0.00536)

When the percentage of null values is superior or equal to 99.50%, it means that the $VaR_{99.5\%}$ for this run is null.

In the table above, we can see that:

- The maximum is 99.5036%, which means that for one of the 10 000 runs, $VaR_{99.5\%} = 0$.
- The minimum is 99.4258%, which means that the percentage of null values for each of the 10 000 runs will never be below 99.4258%.
- On average, the percentage of null values for the 10 000 runs is 99.4656%.

According to the 1st observation, this demonstrate that with a frequency set at 0.00536 there is a chance that $VaR_{99.5\%} = 0$.

We therefore increase little by little the frequency to set a threshold that always has a $VaR_{99.5\%} > 0$.

Distribution of the percentage of null values value in each MC run for $\lambda = 0.00545$		
Minimum	**Max**	**Mean**
99,4200%	99,4966%	99,4565%

Table 7 Distribution of the null values in each MC run (0.00545)

[12] Pseudorandom Number Generator also known as a Deterministic Random Bit Generator (DRBG) is an algorithm for generating sequence of numbers whose properties are similar to sequences of random numbers. In R, the period of return is $2^{31} - 1$.

Distribution of the percentage of null values value in each MC run for $\lambda = 0.00550$		
Minimum	**Max**	**Mean**
99,4144%	99,4874%	99,4514%

Table 8 Distribution of the null values in each MC run (0.00550)

There is no null value in each run for both frequencies. We have in the worst case 99.4966% of the value null, which means that the STEC, which is computed from the $VaR_{99.5\%}$ is different from 0 (99.5%>99.4966%).

We would like to understand if the seed chosen has a real impact on the $VaR_{99.5\%}$. We simulated 1000 times the $VaR_{99.5\%}$ with two frequencies and the same seed sample. We compared the number of 0 that appear in these 1000 values. The seed is the same for each run of the two $VaR_{99.5\%}$.

λ	x	VaR > 0
0.00545	99.466429%	100%
0.00550	99.465434%	100%

Table 9 1000 runs with two frequencies (0.00545 and 0.0055)

Distribution of the percentage of null values value in each MC run for $\lambda = 0.00545$		
Minimum	**Max**	**Mean**
99.4224%	99.4906%	99.4563%

Table 10 Distribution of the null values in each MC run (0.00545)

Distribution of the percentage of null values value in each MC run for $\lambda = 0.0055$		
Minimum	**Max**	**Mean**
99.4174%	99.4864%	99.4513%

Table 11 Distribution of the null values in each MC run (0.0055)

In the table above in the left, we can see that:

- The minimum is 99.4224%, which means that for one of the 1000 runs, we have $quantile(99.4224\%, X_1, \dots, X_n) = 0$, which means that the $VaR(99.5\%)$ is not null.
- The maximum is 99.4906%, which means that for one of the 1000 runs, we have $quantile(99.4906\%, X_1, \dots, X_n) = 0$, which means that the $VaR(99.5\%)$ is not null.
- On average, the null quantile for the 1000 runs is the $quantile(99.4563\% X_1, \dots, X_n)$.

Here, the maximum is obtained with the same seed (934) for both runs. In the worst case (with the maximum), we have a higher null quantile with $\lambda = 0.00545$ $(here,\ quantile(99.4906\%, X_1, \dots, X_n) = 0)$, than with $\lambda = 0.00550$ $(here,\ quantile(99.4864\%, X_1, \dots, X_n) = 0)$, which is logical since $99.4864\% < 99.4906\%$.

We see that the seed have a little impact on the $VaR_{99.5\%}$ with low frequency scenario. For the same seed, we see that we have more positive $VaR_{99.5\%}$ with a higher frequency. The influence of the seed only comes into account when a "bad" seed is chosen, i.e. a seed for which $VaR_{99.5\%} = 0$. That is why we took a preventive frequency threshold found with random seeds.

We have now a frequency threshold. We can set our threshold at $\lambda = 0.00550$. In fact, we know that with $\lambda = 0.00545$ the $VaR_{99.5\%}$ have a low probability to be null. Taking a little higher frequency gives us more conservativeness and the difference between these two values is very low:

$$\Delta\lambda = 0.00005$$

c) Impact of the severity standard deviation

In this section, we want to determine whether or not there is an impact coming from the severity distribution parameters. For that, we simulated 100 times the $VaR_{99.5\%}$ with a fixed seed (one seed for one run), and 100 times the $VaR_{99.5\%}$ with the same seed and a different standard deviation for the lognormal distribution. The seeds chosen here are the seed(100+1) to seed(100+100). We used 500k simulations in each run. We changed the standard deviation of the second $VaR_{99.5\%}$ each of time with a 1/50 step. Here we have taken $\lambda = 0.00545$ and $\sigma = 0.3693$ and a step $k = 0.02$ and:

- a represents the proportion of values superior to 0 with a fixed standard deviation
- b represents the proportion of values superior to 0 with an increasing standard deviation

Run number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Seed	101	102	103	104	105	106	107	108	109	110
a	99,4822%	99,4464%	99,4680%	99,4362%	99,4524%	99,4646%	99,4586%	99,4526%	99,4596%	99,4372%
b	99,4822%	99,4464%	99,4680%	99,4362%	99,4524%	99,4646%	99,4586%	99,4526%	99,4596%	99,4372%
	11	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
Seed	111	112	113	114	115	116	117	118	119	120
a	99,4444%	99,4594%	99,4334%	99,4596%	99,4516%	99,4766%	99,4418%	99,4398%	99,4608%	99,4470%
b	99,4444%	99,4594%	99,4334%	99,4596%	99,4516%	99,4766%	99,4418%	99,4398%	99,4608%	99,4470%
	21	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
Seed	121	122	123	124	125	126	127	128	129	130
a	99,4604%	99,4486%	99,4488%	99,4480%	99,4502%	99,4574%	99,4650%	99,4470%	99,4560%	99,4434%
b	99,4604%	99,4486%	99,4488%	99,4480%	99,4502%	99,4574%	99,4650%	99,4470%	99,4560%	99,4434%
	31	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**
Seed	131	132	133	134	135	136	137	138	139	140
a	99,4602%	99,4518%	99,4544%	99,4408%	99,4522%	99,4376%	99,4482%	99,4454%	99,4526%	99,4452%
b	99,4602%	99,4518%	99,4544%	99,4408%	99,4522%	99,4376%	99,4482%	99,4454%	99,4526%	99,4452%
	41	**42**	**43**	**44**	**45**	**46**	**47**	**48**	**49**	**50**
Seed	141	142	143	144	145	146	147	148	149	150
a	99,4602%	99,4512%	99,4690%	99,4306%	99,4504%	99,4498%	99,4638%	99,4586%	99,4448%	99,4342%
b	99,4602%	99,4512%	99,4690%	99,4306%	99,4504%	99,4498%	99,4638%	99,4586%	99,4448%	99,4342%
	51	**52**	**53**	**54**	**55**	**56**	**57**	**58**	**59**	**60**

Seed	151	152	153	154	155	156	157	158	159	160
a	99,4462%	99,4388%	99,4662%	99,4516%	99,4618%	99,4484%	99,4470%	99,4354%	99,4500%	99,4648%
b	99,4462%	99,4388%	99,4662%	99,4516%	99,4618%	99,4484%	99,4470%	99,4354%	99,4500%	99,4648%
	61	**62**	**63**	**64**	**65**	**66**	**67**	**68**	**69**	**70**
Seed	161	162	163	164	165	166	167	168	169	170
a	99,4578%	99,4436%	99,4626%	99,4578%	99,4362%	99,4410%	99,4686%	99,4622%	99,4686%	99,4506%
b	99,4578%	99,4436%	99,4626%	99,4578%	99,4362%	99,4410%	99,4686%	99,4622%	99,4686%	99,4506%
	71	**72**	**73**	**74**	**75**	**76**	**77**	**78**	**79**	**80**
Seed	171	172	173	174	175	176	177	178	179	180
a	99,4430%	99,4530%	99,4432%	99,4464%	99,4394%	99,4610%	99,4420%	99,4356%	99,4348%	99,4516%
b	99,4430%	99,4530%	99,4432%	99,4464%	99,4394%	99,4610%	99,4420%	99,4356%	99,4348%	99,4516%
	81	**82**	**83**	**84**	**85**	**86**	**87**	**88**	**89**	**90**
Seed	181	182	183	184	185	186	187	188	189	190
a	99,4400%	99,4590%	99,4590%	99,4564%	99,4322%	99,4430%	99,4408%	99,4584%	99,4446%	99,4432%
b	99,4400%	99,4590%	99,4590%	99,4564%	99,4322%	99,4430%	99,4408%	99,4584%	99,4446%	99,4432%
	91	**92**	**93**	**94**	**95**	**96**	**97**	**98**	**99**	**100**
Seed	191	192	193	194	195	196	197	198	199	200
a	99,4558%	99,4534%	99,4218%	99,4576%	99,4592%	99,4520%	99,4712%	99,4672%	99,4370%	99,4616%
b	99,4558%	99,4534%	99,4218%	99,4576%	99,4592%	99,4520%	99,4712%	99,4672%	99,4370%	99,4616%

Table 12 Dual simulations with the same seed

Let's take the run number 61 as an example. For this run, we see that:

- the seed used is the seed 161 for the two runs
- with $\sigma = 0.3693$, we have 99.4578% of the values in this run that are positive. It means that $quantile(0.994578, X_1, \dots, X_n) = 0$, thus $VaR_{99.5\%} > 0$.
- with $\sigma = 0.3693 + 61 \times 0.02 = 1.5893$, we have 99.4578% of the values in this run that are positive. It means that $quantile(0.994578, X_1, \dots, X_n) = 0$, thus $VaR_{99.5\%} > 0$.

We see that with the same seeds, the proportion of value superior to 0 is the same in the two runs at each time. We know that the seeds were the same for each run of the two different vectors. The variation of the standard deviation does not affect the threshold of null value. Basically, we compared $\sigma = 0.3693$ with 100 others values of σ and we always found the same proportion of positive value in each Monte Carlo run. With a fixed seed, we will have more positive value in a Monte Carlo run with a higher frequency.

d) Impact of the number of simulations

To see the impact of the number of simulations, we run 1000 times the $VaR_{99.5\%}$ for various numbers of simulations. We tested the following numbers: 500k, 750k, 1M, 2M and 5M. We did the computation for two values of frequencies: $\lambda = 0.00545$ and $\lambda = 0.0055$. Also, in order not to have the influence of the seed, we took the same seed for each dual run.

Frequency	Number of simulations	Min	Max	Mean
0.00545	500k	99.4224%	99.4906%	99.4563%
0.0055	500k	99.4174%	99.4864%	99.4513%
0.00545	750k	99.4312%	99.4844%	99.4565%
0.0055	750k	99.4268%	99.4798%	99.4516%
0.00545	1M	99.4307%	99.4823%	99.4565%
0.0055	1M	99.4254%	99.4769%	99.4516%
0.00545	2M	99.4387%	99.4749%	99.4566%
0.0055	2M	99.4338%	99.4694%	99.4516%
0.00545	5M	99.4452%	99.4686%	99.4564%
0.0055	5M	99.4403%	99.4636%	99.4514%

Table 13 Monte Carlo run with various numbers of simulations

Each time the number of simulations increases, the proportion of positive values increases too. We can conclude that the number of simulations has an impact on the frequency threshold.

e) Conclusion

We have found a theoretical threshold of $\lambda = 0.005012$. With the studies that have been done we found that the seed have a little impact on the threshold. To offset the impact of the seed, we take a preventive frequency threshold by setting it to a “higher” level. The standard deviation of the lognormal distribution does not affect the threshold under which the related quantile to the frequency is null. It is not surprising because of independence hypothesis between the frequency and the severity distribution. With the frequency parameter λ, the more impacting parameter is the number of simulations in the Monte Carlo runs. In conclusion, the value to $\lambda = 0.0055$ chosen in section 3 is an adequate threshold. We are confident that $VaR_{99.5\%} > 0$ if we take a frequency superior to this value.

Chapter 5: Aggregation approach

After having calibrated all the scenarios, experts build correlations between the risks to aggregate them. The needed properties of the correlation matrix will be presented on this master thesis. The most common aggregation approaches used in operational risk are:

- Variance-covariance approach
- Copula approach

Also, we have to bear in mind that a diversification effect is introduced with the correlation matrices. The diversification represents a fundamental concept adopted in almost all areas of the insurance industry. It is not just in the insurance industry that the diversification is important, but also in the banking and finance industry. The diversification can be introduced in various levels:

- between risk scenarios: within each category of operational risk in each entity
- Entity level: correlation between the entities or also called geographical correlation

To see the effect of the diversification on the Operational Risk SCR, studies on the study of the geographical correlation and the correlation between operational risk categories are presented in this chapter.

1. Correlation matrix

In a SBA, the correlation matrices are built by experts when assessing the scenarios and the dependencies between them. Certain properties are not met when constructing these correlation matrices. In this section we will study the needed properties of the correlation matrices, as they are essential in the modeling of the operational risk, especially in the aggregation.

The dependence is in statistics the relationship between two random variables. It refers to all the random variables that do not respect the following property:

$$(X_1, \dots, X_n) \ are \ independents \ if \ and \ only \ if:$$

$$\forall (x_1, \dots, x_n) \in \mathbb{R}^n, \mathbb{P}\left(\bigcap_{i=1}^{n} \{X_i \leq x_i\}\right) = \prod_{i=1}^{n} \mathbb{P}(X_i \leq x_i)$$

We consider the correlation as a measure of dependence. The most used correlations are: linear correlation, rank correlation and tail dependence.

a) Linear correlation

It is the most used and popular measure of correlation. It is called the Pearson correlation coefficient. Let's denote (X, Y) two random variables. The Pearson correlation coefficient is defined as follows:

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = \frac{\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)}{\sigma_X \sigma_Y}$$

$$where\ \sigma_X, \sigma_Y\ are\ positive\ and\ the\ standard\ deviation\ of\ X\ and\ Y.$$

We can use Pearson correlation coefficient to build a correlation matrix as:

- It is a measure of linear dependence
- It is symmetric $\rho(X, Y) = \rho(Y, X)$
- $\forall (X, Y)\ random\ variables: -1 \leq \rho(X, Y) \leq 1$
- It is invariant to linear transformation on the variables

The case where $\rho(X, Y)$[13] $= 0$ does not imply that the variables (X, Y) are independents.

If the following property is met, then X and Y are independents:

$$Cov\big(f(X), g(Y)\big) = 0, for\ any\ f\ and\ g\ function.$$

An example of two dependent random variables with a null covariance is easy to find. If X is a symmetric and centered distribution with a null mean, then X and X^2 have a null covariance but are clearly dependent. An application is with the $\mathcal{N}(0,1)$ distribution. If $X \sim \mathcal{N}(0,1)$:

$$Cov(X, X^2) = \mathbb{E}(XX^2) - \mathbb{E}(X)\mathbb{E}(X^2), \qquad but\ \mathbb{E}(X) = 0\ and\ \mathbb{E}(X^3) = 0$$

$$then\ Cov(X, X^2) = 0$$

In a general manner, we have that all the impair moments of the $\mathcal{N}(0,1)$ are null.

b) Rank correlation

A rank correlation is a statistic that measures the relationship between two variables by measuring the degree of similarity orderings of the set of data. It is an alternative to the linear correlation; we use the potential dependencies existing in the orderings of the data.

The two most popular rank correlations are:

[13] $\rho(X, Y) = 0 \Leftrightarrow Cov(X, Y) = 0$

- Spearman's rho
- Kendall's tau

<u>Spearman's rho</u>

The Spearman correlation coefficient is defined as the Pearson correlation coefficient between the ranks of set of data of the random variables. Let's denote $X, Y \in \mathbb{R}^n$ two vectors of n simulations and $X^R, Y^R \in \mathbb{R}^n$ the ranked simulations of those vectors. The Spearman correlation coefficient is defined by:

$$\widehat{\rho_S} = 1 - \frac{6\sum_{i=1}^{n} {d_i}^2}{n^3 - n}, \qquad where\ d_i = X_i^R - Y_i^R$$

One must have special attention on the cases where there are a lot of ties in the sample. As the Spearman's rho is a special case of the Pearson's coefficient, we have the same properties:

- Non-parametric measure
- It is symmetric $\widehat{\rho_S}(X, Y) = \widehat{\rho_S}(Y, X)$
- $\forall (X, Y)\ random\ variables\colon\ -1 \leq \widehat{\rho_S}(X, Y) \leq 1$

<u>Kendall's tau</u>

The Kendall's tau coefficient is as statistic to measure the association between two variables. It uses the number of concordant and discordant pairs:

- A pair of observations i, j is considered as concordant if and only if:

$$(x_i > x_j\ and\ y_i > y_j)\ or\ (x_i < x_j\ and\ y_i < y_j)\ \ and\ C = \sum concordant\ pairs$$

- A pair of observations i, j is considered as discordant if and only if:

$$(x_i < x_j\ and\ y_i > y_j)\ or\ (x_i > x_j\ and\ y_i < y_j)\ \ and\ D = \sum discordant\ pairs$$

Let's denote n the size of the two simulated vectors X and Y. The Kendall's tau correlation coefficient is defined as following:

$$\hat{\tau} = \frac{C - D}{\frac{1}{2}n(n-1)}$$ [14]

[14] $\frac{1}{2}n(n-1) = \binom{n}{2}$, the number of pairs.

2. Study of the Operational Risk categories correlation

In AXA aggregation approach, a diversification between the operational risk categories needs to be defined. Indeed, the diversification benefit is justified by two main reasons:

- It is unlikely to consider that all severe operational risk losses (e.g. an internal fraud in Korea & an external fraud in France) occurs simultaneously & systematically in the same year across the Group
- A level of diversification between the operational risk categories and entities is therefore introduced to reflect this reality into the internal model calculation

Generally, the level of correlation between operational risk categories is considered to be low given the nature of the risks evaluated:

- Considering the nature of the major operational risks a simple additive aggregation is not appropriate to reflect the risk profile of the company

The aim of this section is to demonstrate the low level of correlation between the operational risk categories.

We choose here to use the Kendall's rank correlation for the following reasons:

- The distribution of Kendall's tau has better statistical properties than the distribution of Spearman's rank correlation
- The interpretation of Kendall's tau in terms of probabilities of observing the agreeable (concordant) and non-agreeable (discordant) pairs is very direct
- In most of the situation, the interpretations of Kendall's tau and Spearman's rank correlation coefficient are very similar and thus invariably lead to the same inferences

For each pair of entities we have:

- Computed the Kendall's rank correlation
- Rounded down the above computation to the nearest 25% (we will then have between -100% and 100% with a 25% step for the correlation)
- Built the distribution of the correlation between entities among the possible values (multiple of 25%) Built the distribution of the correlation between entities among the possible values (multiple of 25%)

In a publication of Frachot, Roncalli and Salomon in 2004[15], the correlations between aggregate losses may results from:

[15] FRACHOT A., RONCALLI T. and SALOMON E., "The Correlation Problem in Operational Risk", Groupe de Recherche Opérationnelle du Crédit Agricole, France, 2004

- Correlation between frequencies
- Correlation between severities
- Correlation between both

AXA’s approach is to consider the aggregated loss correlation conveyed by underlying correlations between frequencies, i.e. the number of losses.

We want to demonstrate the low correlation between Operational Risk. We based our study on the internal losses reported by each entities and studied the dependencies between the operational risk categories. We choose to take the following scope of time: 2008-2014. As we have here operational risk categories, we don’t have to treat the data anymore. We remind that there are seven operational risk categories:

- Internal Fraud
- External Fraud
- Employment Practices and Workplace Safety
- Clients, Products and Business Practices
- Damage to Physical Assets
- Business disruption and system failures
- Execution, Delivery and Process Management

a) Correlation on treated data

The first correlations computed are as in the previous section the Kendall’s correlation by number of losses of the operational risk categories. Instead of comparing the treated data and untreated data, we computed the correlations by number of losses on a yearly and quarterly basis.

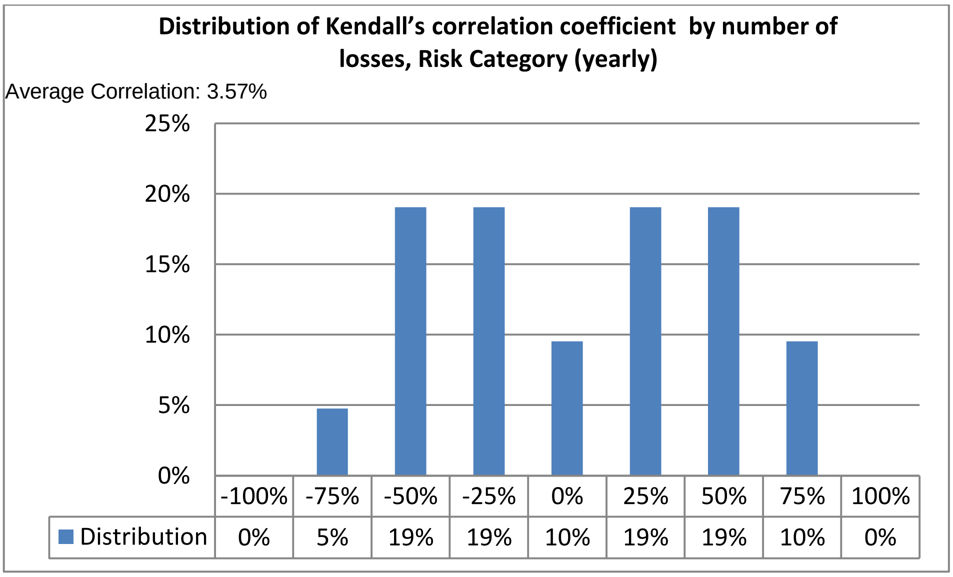
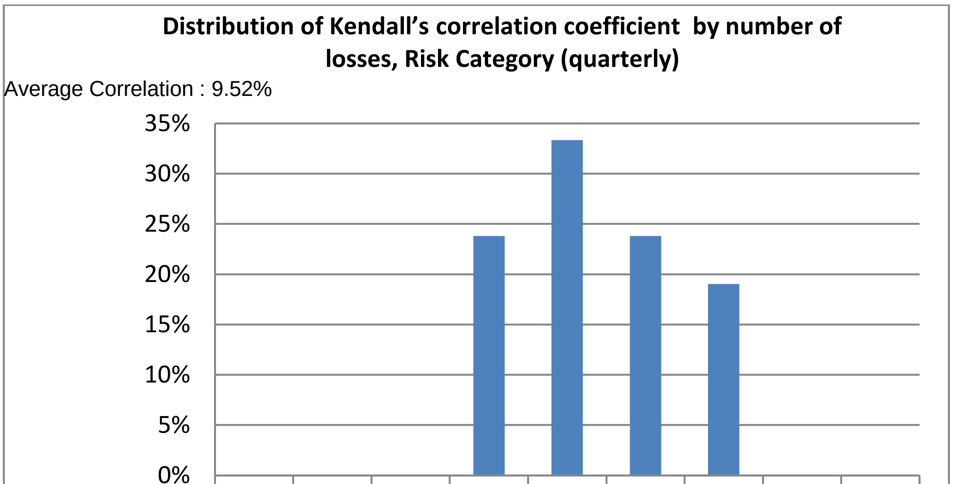


Figure 4 Frequency Correlation between Operational Risk categories Yearly



	-100%	-75%	-50%	-25%	0%	25%	50%	75%	100%
Distribution	0%	0%	0%	24%	33%	24%	19%	0%	0%

Figure 5 Correlations between Operational Risk categories Quarterly

The main reason for computing the quarterly Frequency correlation is to see how the distribution of the correlation coefficients behaves with a bigger sample. The yearly correlations seem to have a bigger standard deviation than the quarterly correlations. Although computing correlations on quarterly data seems more satisfying as the volume of data is higher and the repartition of the correlation seems symmetric, it is not adequate as we need to model dependencies between annual losses.

To see if this value is adequate or not, we decided to compute the correlation between entities with various approaches:

- Severity correlation: the maximum loss amount by year
- Aggregate loss correlation: the total losses amount of a year
- Frequency / Severity correlation: the average amount of losses by year

We compare then the three above correlation distributions with the frequencies correlations computed.

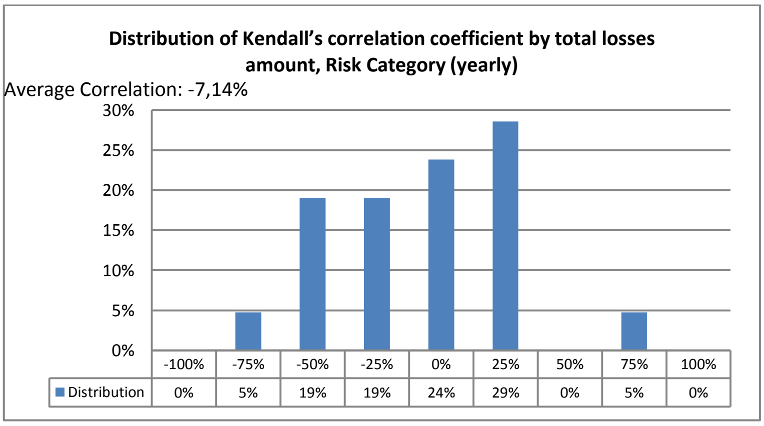


Figure 6 Kendall's correlation by total losses amount, Risk Category

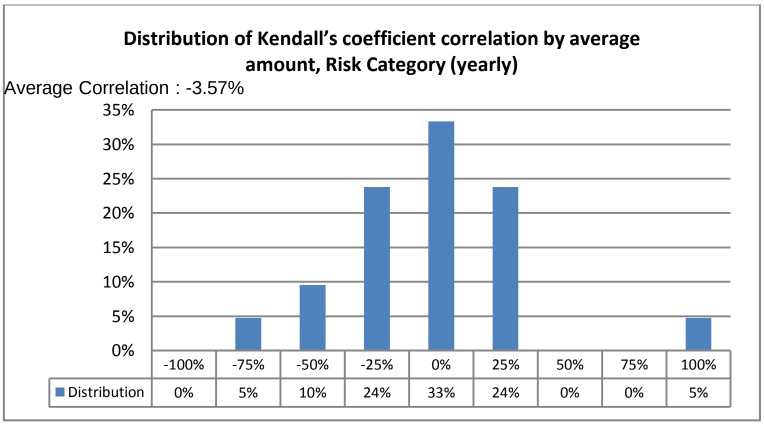


Figure 7 Kendall's correlation by average amount, Risk Category

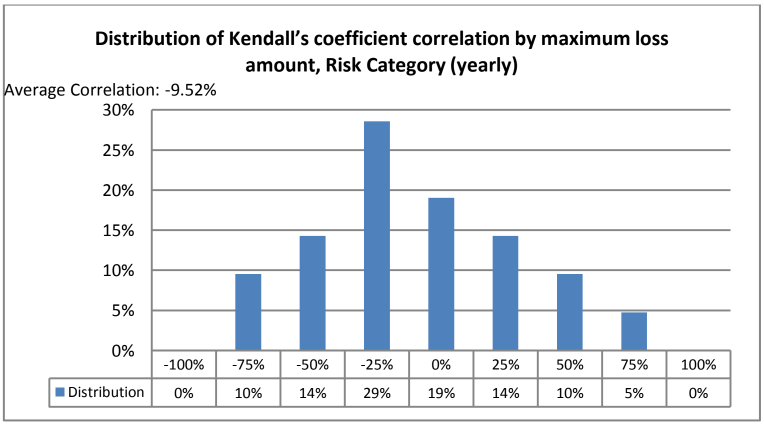


Figure 8 Kendall's correlation by maximum loss amount, Risk Category

We see that the Frequency Correlation approach is the more conservative among the four types of correlations computed. The correlations computed on loss amount data are on average between -10% and -3%.

Due to the lack of data for annual number of losses, a Bootstrap method is used to better understand the correlations.

b) Application of the bootstrap

<u>Definition: Bootstrap</u>

> Bootstrapping can refer to any test or metric that relies on random resampling with replacement. In practice, bootstrapping is used for estimating an estimator's properties (such as the mean or variance).

In order to apply the bootstrap to resample the data, the approach that has been used will be presented.

Let's denote the $E = (E_1, \dots, E_n)$ the matrix of the entities losses with $E_1 = \begin{pmatrix} E_1(1) \\ \vdots \\ E_1(m) \end{pmatrix}$ where $E_1(j), \forall j \in [\![1, m]\!]$ is the annual number of losses of the year j.

We have $E(j) = (E_1(j), \dots, E_n(j))$, $\forall j \in [\![1, m]\!]$, thus we can rewrite the matrix $E = \begin{pmatrix} E(1) \\ \vdots \\ E(m) \end{pmatrix}$. We will apply the bootstrap approach by creating new sample of $E^b = \begin{pmatrix} E^b(1) \\ \vdots \\ E^b(m) \end{pmatrix}$, with $E^b(j) = E(i), \forall (i,j) \in [\![1, m]\!]^2$ where $E(i)$ is taken from $\begin{pmatrix} E(1) \\ \vdots \\ E(m) \end{pmatrix}$ with a uniform discrete distribution $\mathcal{U}\{1, m\}$. The approach used is:

- We fix a number of iteration B (usually, $B = 1\ 000$), and we create B bootstrap samples:

$$\forall b \in [\![1, B]\!], E^b = \begin{pmatrix} E^b(1) \\ \vdots \\ E^b(m) \end{pmatrix} = \begin{pmatrix} E_1^b(1) & \dots & E_n^b(1) \\ \vdots & \ddots & \vdots \\ E_1^b(m) & \dots & E_n^b(m) \end{pmatrix}$$

- We compute the correlation matrix by computing the Kendall's correlation coefficient between each vector $(E_1^b, \dots, E_n^b)$ with $\tau(.,.)$ the function that measure the Kendall's rank correlation:

$$C^b = \left(\tau\left(E_i^b, E_j^b\right)\right)_{i,j}, \forall (i,j) \in [\![1, n]\!]^2$$

- We take the mean of the B iterations. And we obtain a matrix:

$$\bar{C} = \left(\overline{\tau\left(E_i^b, E_j^b\right)}\right)_{i,j}, with\ \overline{\tau\left(E_i^b, E_j^b\right)} = \frac{1}{B}\sum_{b=1}^{B} \tau\left(E_i^b, E_j^b\right), \forall (i,j) \in [\![1, n]\!]^2$$

We obtain the following distribution of the correlation coefficient:

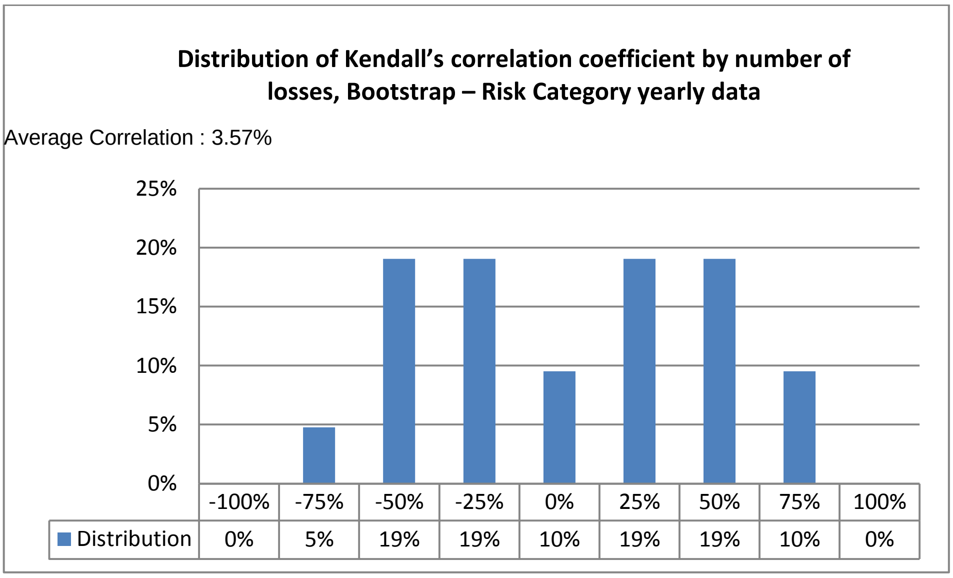


Figure 9 Frequency Correlation between Operational Risk categories Yearly - Bootstrap

As we can see, the results of the bootstrap approach give the same results due to the rounding to the nearest 25%.

But we want to know if the number of iteration in the bootstrap approach is sufficient or not. According to TAHANI A. MATURI's paper research in 2010[16]:

$$\mathbb{P}\left(100 \times \frac{|\tau^B - \tau^\infty|}{\tau^\infty} \leq PerD\right) = 1 - \alpha$$

where τ^∞ is the "ideal" bootstrap estimate, τ^B is the bootstrap approximation of τ^∞ after B iterations, $PerD$ the Percentage of Deviation and α the level of error.

Here we have taken $B = 1\,000$, we know that for $B = 1\,029$ we have: $PerD = 5\%$ and $\alpha = 5\%$.

If we want to optimize the Percentage of Deviation and the level of error, we have to take:

- $B = 1\,839 \; for \; PerD = 5\% \; and \; \alpha = 1\%$
- $B = 20\,843 \; for \; PerD = 1\% \; and \; \alpha = 5\%$
- $B = 39\,455 \; for \; PerD = 1\% \; and \; \alpha = 1\%$

[16] TAHANI A. MATURI, "A Comparison of Correlation Coefficients via a Three-Step Bootstrap Approach", Journal of the Mathematics research Vol. 2, No 2, 2010.

3. Geographical correlation analysis

The study on the operational risk categories correlation was done with the same methodology used for the correlation between operational risk categories. We want to demonstrate the low level of the geographical correlation.

a) Kendall's rank correlation on untreated and treated data

The first approach is to compute the correlations between entities with the data untreated from 2008 to 2014. The high heterogeneity of the data before 2008 does not allows us to include these data. We computed the first correlations with a scope of time of 6 years. More than 35 000 losses are studied for over 50 entities to compute the geographical correlation matrix. For the first correlation matrix, the correlation between frequencies has been selected. The method to compute the correlation matrix will be presented in the next section (b).

The results are the following:

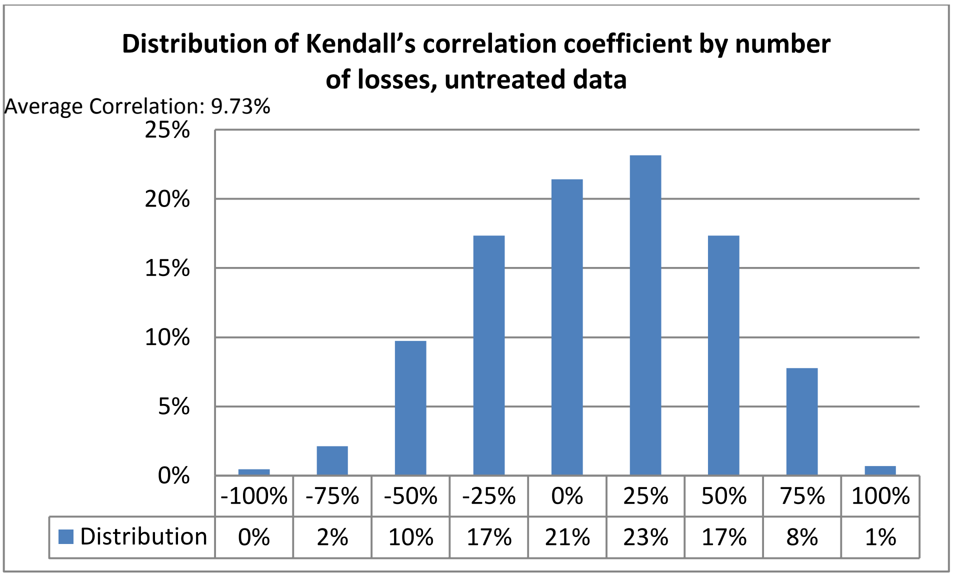


Figure 10 Distribution of Kendall correlation, untreated data

In average, we have 9.73% of correlation between two entities. Also, we have around 50% of the correlation coefficient that are negative or null. It clearly does not reflect the dependence structure of the entities in practice. But given the scarcity and heterogeneity of the data, we have to treat them and select the entities that have consistent and coherent data.

We selected a fewer number of entities in order to study the geographical correlation. We kept the same scope of time: 2008-2014.

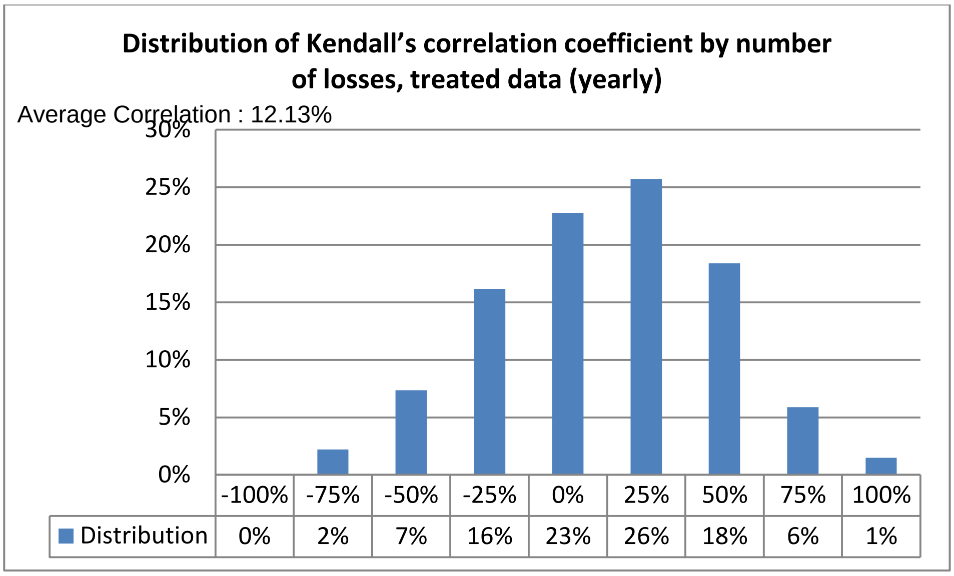


Figure 11 Distribution of Kendall's correlation, treated data

Here, we have an average correlation of 12.13%. To see if this value is adequate or not, we decided to compute the correlation between entities with various approaches:

- Severity correlation: the maximum loss amount by year
- Aggregate loss correlation: the total losses amount of a year
- Frequency / Severity correlation: the average amount of losses by year

We compare then the three above approaches with the frequencies correlations computed.

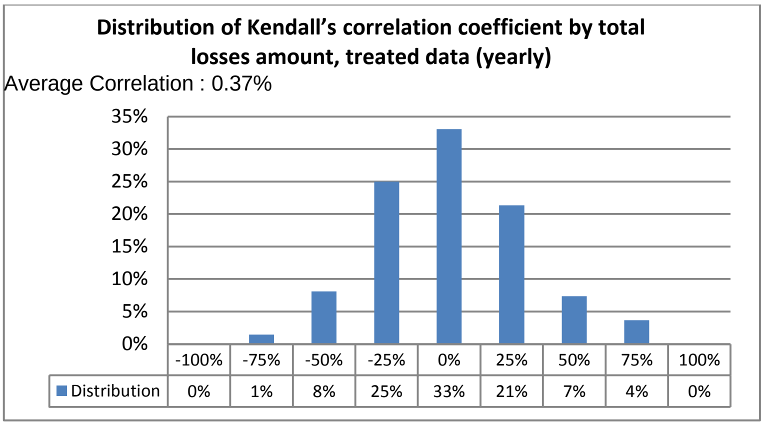


Figure 12 Kendall's correlation by total losses amount

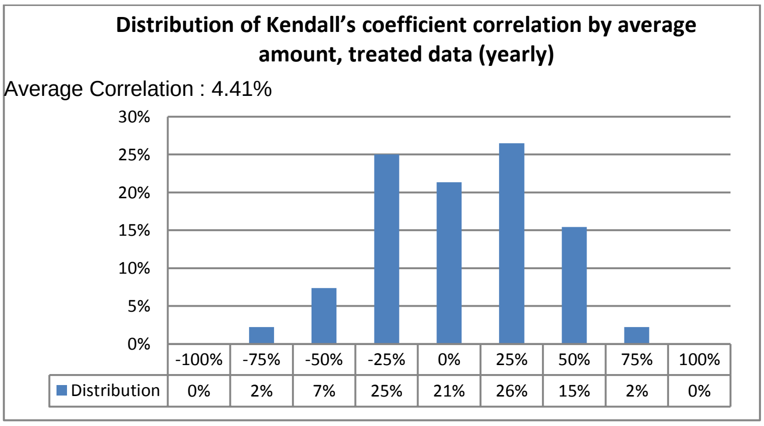


Figure 13 Kendall's correlation by average amount

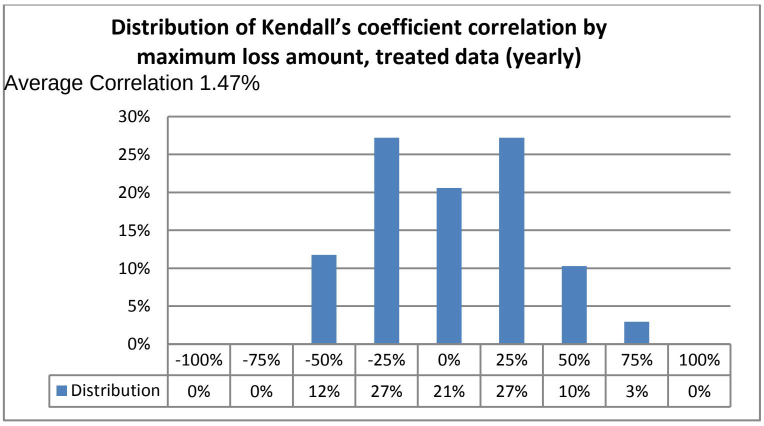


Figure 14 Kendall's correlation by maximum loss amount

We see that the correlations computed between frequencies are the more conservative. The average of the correlations does not vary too much for these three types of correlation computed: it is between 0% and 5%.

Due to the lack of data for annual number of losses, a Bootstrap method is used to better understand the correlations.

b) Application of the Bootstrap

As a last comparison, we computed the distribution of the correlation with the same bootstrap approach used in the operational risk categories correlation study. We choose to apply the bootstrap to the Frequency Correlation.

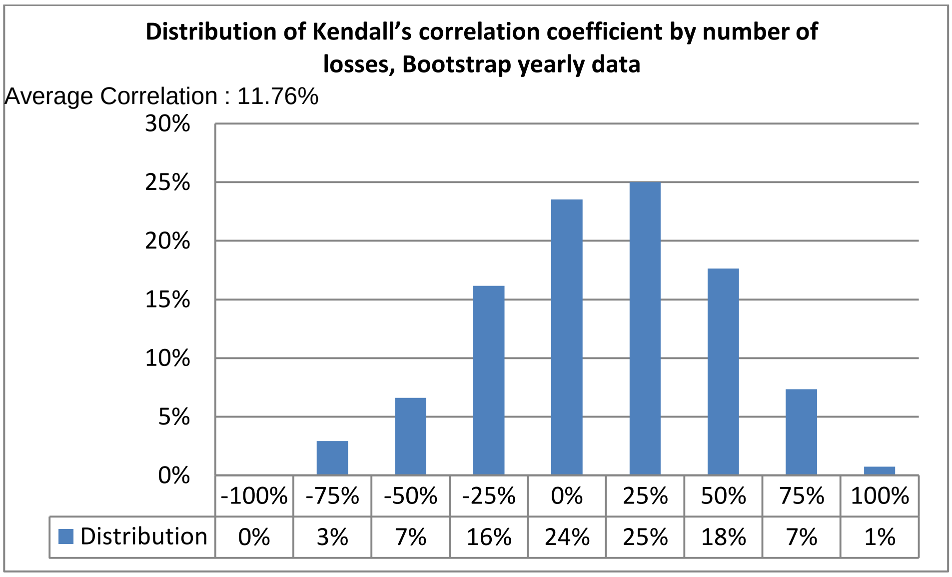


Figure 15 Kendall's correlation Bootstrap

We see that the average is still around 12% with $B = 1\,000$ iterations. We have a demonstrated that we have a low geographical correlations level.

4. Correlation matrix needed property

Now that we have defined the measure of dependence for our risks, we are able to define a correlation matrix in order to aggregate the risks. When simulating a multivariate $\mathcal{N}_n(0_n, \Sigma)$ normal distribution, the correlation matrix Σ must be positive definite (PD) in order to have a non-degenerated distribution and to perform the Cholesky decomposition algorithm on Σ.

<u>Definition: Positive Definite matrix</u>

Let's denote

$M \in \mathcal{M}_n(\mathbb{R}) = \{set\ of\ squarred\ symmetrical\ matrix\ of\ dimension\ n \times n\}$,

M is considered as a Positive Definite matrix if one of the following properties is met:

- For any $\forall u \in \mathbb{R}^n, with\ u \neq 0_n$ we have: $u^t M u > 0$
- All the eigenvalues of M are positive (> 0)
- The symmetric bilinear form $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, (x, y) \longmapsto x^t M y$ is a scalar product in $\mathbb{R}^n$
- M has a unique Cholesky decomposition

- All the leading principal minors[17] are positive (> 0)

<u>Cholesky's theorem:</u>

If M is a symmetric positive definite matrix, then, there exists a unique triangular inferior matrix L such that:

$$M = LL^t$$

And all the diagonal elements of L are positive (> 0).

To compute the Cholesky decomposition matrix, the following algorithm can be applied. We want to obtain a matrix:

$$L = \begin{pmatrix} l_{11} & & & \\ l_{21} & l_{22} & & \\ \vdots & \vdots & \ddots & \\ l_{n1} & l_{n2} & \dots & l_{nn} \end{pmatrix} with\ M = LL^t\ a\ M\ is\ symmetric\ positive\ definite$$

With $M = LL^t$ we can deduce that

$$m_{ij} = (LL^t)_{ij} = \sum_{k=1}^{n} l_{ik}\, l_{jk} = \sum_{k=1}^{min\{i,j\}^{18}} l_{ik}\, l_{jk}, \forall (i,j) \in [\![1,n]\!]^2$$

Since M is symmetric, we only need that the above relation is met for $i \leq j$, i.e. we want L to satisfy $m_{ij} = \sum_{k=1}^{i} l_{ik}\, l_{jk},\ 1 \leq i \leq j \leq n$:

- For $i = 1$, we compute the first column of L

$$m_{11} = l_{11}l_{11} \ \Rightarrow\ l_{11} = \sqrt{m_{11}}\ and\ m_{1j} = l_{11}l_{j1} \ \Rightarrow l_{j1} = \frac{m_{1j}}{l_{11}}, 2 \leq i \leq n$$

- The $i - th$ column of L $(2 \leq i \leq n)$ is computed with the first $(i-1) - th$ columns

$$m_{ii} = l_{i1}l_{i1} + \cdots + l_{ii}l_{ii} \ \Rightarrow\ l_{ii} = \sqrt{m_{ii} - \sum_{k=1}^{i-1} (l_{ik})^2}$$

$$m_{ij} = l_{i1}l_{j1} + \cdots + l_{ii}l_{ji} \ \Rightarrow\ l_{ji} = \frac{m_{ij} - \sum_{k=1}^{i-1} l_{ik}l_{jk}}{l_{ii}}, \qquad i+1 \leq j \leq n$$

There is a case where the Cholesky decomposition is not unique. It is when the matrix M is positive semi-definite.

---

[17] The $k - th$ leading principal minor of M is the determinant of its upper-left k by k sub-matrix. M is positive definite if and only if all these determinant are non-negative (> 0). It is called the Sylvester's criterion.

[18] $l_{ij} = 0\ if\ 1 \leq i < j \leq n$

Also, we have that:

> A Cholesky decomposition is unique if and only if all the diagonal elements of the triangular matrix are positive.

If we have a Cholesky decomposition with a 0 as a diagonal element, the matrix is positive semi-definite and not unique.

<u>Definition: Positive Semi-Definite matrix</u>

Let's denote

$M \in \mathcal{M}_n(\mathbb{R}) = \{set\ of\ squarred\ symmetrical\ matrix\ of\ dimension\ n \times n\}$,

M is considered as a Positive Semi-Definite matrix if one of the following properties is met:

- For any $\forall u \in \mathbb{R}^n, with\ u \neq 0_n$ we have: $u^t M u \geq 0$
- All the eigenvalues of M are non-negative(≥ 0)
- All the leading principal minors are non-negative (≥ 0)

If a matrix is positive semi-definite, then it results on having a null diagonal element in the Cholesky decomposition matrix. In AXA's approach, only the PD matrices are used in the aggregation approach.

5. Nearest correlation matrix

When we build with the experts the correlation matrix, the positive definiteness property is not always met. For this reason, one could try to change the correlation coefficient. To avoid changing the expert's opinion on the correlation coefficient, two methods were developed in order to compute the nearest positive definite correlation matrix.

a) Rebonato's algorithm

This approach was developed by Rebonato and Jäcker in 1999[19]. We will present the algorithm to compute the nearest correlation matrix. It is a method called, spectral decomposition or principal component analysis: we retrieve a positive definite matrix from eigenvalues and eigenvectors.

[19] REBONATO AND JÄCKEL, "The most general methodology to create a valid correlation matrix for risk management and option pricing purposes", 1999.

Rebonato's algorithm on a squared symmetric matrix $M \in \mathcal{M}_n(\mathbb{R})$:

- Compute the eigenvalues and eigenvectors of M, and define

$$\sigma(M) = \{eigenvalues\ of\ M\} = \{\lambda_1, \dots \lambda_n\}, and\ \Lambda = diag(\sigma(M))$$

- All the eigenvalues smaller than a threshold ε[20] (> 0) will be replaced by ε

$$\Lambda' = \{\lambda_1', \dots \lambda_n'\}, \qquad with\ \lambda_i' = \max(\lambda_i, \varepsilon), \forall i \in [\![1, n]\!]$$

- We define now a scaling matrix $T = (t_i)_{i=1,\dots,n}$ such that:

$$\forall i \in [\![1, n]\!], \qquad t_i = \frac{1}{\sum_{k=1}^{n} (v_{i,k})^2 \lambda_k'}\ where\ v\ is\ the\ matrix\ of\ the\ eigenvectors\ of\ M$$

- Next is defining the matrix $B' = v\sqrt{\Lambda'}$ and to normalize it with T to have a slightly different matrix that is positive definite: $B = \sqrt{T}B'$=$\sqrt{T}\,v\sqrt{\Lambda'}$, and we have $BB^t = \sqrt{T}\,v\sqrt{\Lambda'}\left(\sqrt{T}\,v\sqrt{\Lambda'}\right)^t = \sqrt{T}\,v\sqrt{\Lambda'}\left(\sqrt{\Lambda'}\right)^t v^t\left(\sqrt{T}\right)^t = v'\Lambda'(v')^t$ with $v' = \sqrt{T}\,v$ the matrix of the eigenvectors normalized.
- The matrix $\bar{M} = BB^t$ is the nearest positive definite correlation matrix.

b) Higham's algorithm

Nicholas Higham developed this approach in 2002[21]. It is an alternative approach to the one developed by Rebonato and presented in the previous section. It consists on projecting the set of symmetric matrices onto the correlations matrices. The details of the method are available in appendix. Only the algorithm to apply will be presented in this section.

Let's denote a squared symmetric matrix $M \in \mathcal{M}_n(\mathbb{R})$, $\Delta S_0 = 0, Y_0 = M,\ \varepsilon = M, k = 1\ and\ tolerance = 0.0001$. As long as $\|\ \varepsilon\| > tolerance$:

- $R_k = Y_{k-1} - \Delta S_{k-1}$
- $X_k = v_k \Lambda_k (v_k)^t$, with

$$(\Lambda_k)_{i,j} = \begin{cases} \max(tolerance, \lambda_i), if\ i = j \\ 0 \qquad\qquad\qquad\qquad\quad, else \end{cases}, \forall (i,j) \in [\![1, n]\!]^2\ and\ v_k\ the\ eigenvector\ of\ R_k$$

- $\Delta S_k = X_k - R_k$, we set $temp = Y_k$
- $Y_k = X_k - diag(X_k - I_n)$
- $\varepsilon = temp - Y_k$

---

[20] For approximation purpose, we consider a matrix to be positive definite if all its eigenvalues are superior to ε.

[21] Nicholas HIGHAM, "Computing the Nearest Correlation Matrix – A problem from Finance", Manchester Institute for Mathematical Sciences 2002

- $k = k + 1$

The above algorithm as well as the Rebonato's algorithm has been implemented in R and VBA.

c) Comparison of the two algorithms

Let's take as an example the following non-PD matrix:

1	0.75	0.5	0.5	0.5	0.5	0.5	0.5
0.75	1	0.75	0.5	0.5	0.5	0.5	0.5
0.5	0.75	1	0.75	0.5	0.5	0.5	0.5
0.5	0.5	0.75	1	0.75	0.25	0.5	0.5
0.5	0.5	0.5	0.75	1	0.75	0.5	0.5
0.5	0.5	0.5	0.25	0.75	1	0.75	0.5
0.5	0.5	0.5	0.5	0.5	0.75	1	0.75
0.5	0.5	0.5	0.5	0.5	0.5	0.75	1

Table 14 Non-PD correlation matrix

The above matrix is not PD due to one of its eigenvalue being negative (-0.085). We obtain the following matrix when we apply the two algorithms:

1	0.7343	0.4981	0.4868	0.4986	0.4866	0.4992	0.4915
0.7343	1	0.7342	0.4999	0.4872	0.4999	0.4864	0.4991
0.4981	0.7342	1	0.7238	0.5089	0.4791	0.5040	0.4880
0.4868	0.4999	0.7238	1	0.7124	0.2741	0.4807	0.5016
0.4986	0.4872	0.5089	0.7124	1	0.7128	0.5065	0.4906
0.4866	0.4999	0.4791	0.2741	0.7128	1	0.7258	0.4987
0.4992	0.4864	0.5040	0.4807	0.5065	0.7258	1	0.7324
0.4915	0.4991	0.4880	0.5016	0.4906	0.4987	0.7324	1

Table 15 Higham PD Correlation Matrix

1	0.7477	0.5003	0.4895	0.499	0.4897	0.5004	0.4988
0.7477	1	0.7421	0.5	0.4888	0.5001	0.4941	0.5004
0.5003	0.7421	1	0.7237	0.5061	0.478	0.504	0.4952
0.4895	0.5	0.7237	1	0.7127	0.2668	0.4792	0.4978
0.499	0.4888	0.5061	0.7127	1	0.7133	0.5053	0.4914
0.4897	0.5001	0.478	0.2668	0.7133	1	0.7258	0.4979
0.5004	0.4941	0.504	0.4792	0.5053	0.7258	1	0.7447
0.4988	0.5004	0.4952	0.4978	0.4914	0.4979	0.7447	1

Table 16 Rebonato PD Correlation Matrix

We have now two PD matrices. The choice between the two matrices is let to the expert or the risk manager in charge. Below is a comparative table of the capital charge of 8 aggregated scenarios of the two transformed PD matrices with the original matrix. As we cannot use the copula approach due to the non-PD matrix, we used the variance-covariance approach to do the calculation of the capital charge.

	Higham	Rebonato
Difference	-0.3265%	-0.1505%

Table 17 Variation of the aggregated capital charge with respect to the non-PD matrix

6. The different aggregation approaches

a) Variance-covariance approach

This approach is the easiest to implement with a correlation matrix. The aggregated SCR can be computed with the following formula:

$$SCR_{aggr} = \{u^t \Sigma u\}^{\frac{1}{2}}, with\ u = \begin{pmatrix} SCR_{scenario\ 1} \\ \vdots \\ SCR_{scenario\ n} \end{pmatrix}\ and\ \Sigma\ a\ symmetric\ correlation\ matrix$$

$$\Sigma = \begin{pmatrix} 1 & \rho_{1,2} & \cdots & \rho_{1,n} \\ \rho_{2,1} & 1 & \cdots & \rho_{2,n} \\ \vdots & & \ddots & \vdots \\ \rho_{n,1} & \rho_{n,2} & \cdots & 1 \end{pmatrix}, with\ \rho_{i,j} = \rho_{j,i}\ and\ -1 \leq \rho_{i,j} \leq 1\ \forall (i,j) \in [\![1, \dots, n]\!]^2$$

Here $\rho_{i,j}$ is the correlation between the scenario i and j.

It is a close formula, easy to compute. But we need to remind that there must be a Gaussian assumption: the vector of annual losses must follow a Gaussian distribution, i.e. if $S = \begin{pmatrix} S_1 \\ \vdots \\ S_n \end{pmatrix}$ is the vector of annual losses, then:

$$S \sim \mathcal{N}(\mu, \Sigma)$$

$$where\ \mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}\ and\ \Sigma = \begin{pmatrix} 1 & \rho_{1,2} & \cdots & \rho_{1,n} \\ \rho_{2,1} & 1 & \cdots & \rho_{2,n} \\ \vdots & & \ddots & \vdots \\ \rho_{n,1} & \rho_{n,2} & \cdots & 1 \end{pmatrix}, -1 \leq \rho_{i,j} \leq 1\ \forall (i,j) \in [\![1, \dots, n]\!]^2$$

For this approach, we only need the quantified scenario risks, and the correlations. It is the fastest approach in terms of computation time, as it does not require having the scenario simulation trips. Nevertheless, the vector of annual losses must have a Gaussian structure, which is not the case in an operational risk framework. Another drawback of the Variance-Covariance approach is that the (right) tail dependencies are not taken into account, which is the contrary of the Copula approach that will be described in the next section.

b) Copula approach

A common alternative to the variance-covariance approach is the copula approach. It allows catching the dependencies between the operational risks scenarios or entities.

Definition: Copula

A copula is a multivariate probability distribution function for which the marginal probability distributions of each variable are uniform distribution defined on $[0,1]$. Thus, a copula is a function defined on $[0,1]^n$ by:

$$C(u_1, \dots, u_n) = \mathbb{P}(U_1 \leq u_1, \dots, U_n \leq u_n), \qquad \forall (u_1, \dots, u_n) \in [0,1]^n$$

The fundamental result and tool of copulas is Sklar's theorem:

If $F = F_{X_1,\dots,X_n}$ is the cumulative distribution function of the marginal distribution $F_{X_1}, \dots, F_{X_n}$, then there exists an n-copula C such that for any $(x_1, \dots, x_n) \in \mathbb{R}^n$:

$$F(x_1, \dots, x_n) = C\left(F_{X_1}(x_1), \dots, F_{X_n}(x_n)\right)$$

Conversely, if C is an n-copula and $F_{X_1}, \dots, F_{X_n}$ are cumulative distribution functions, then the function $F = F_{X_1,\dots,X_n}$ is an n-dimensional distribution function with marginal functions $F_{X_1}, \dots, F_{X_n}$.

If $F_{X_1}, \dots, F_{X_n}$ are continuous, then C is unique.

We can deduce from this theorem that we are able to represent the dependence structure with a copula.

Various types of copula exist. One of the most simple to implement is the Gaussian copula.

Definition: Gaussian Copula

$$C(u_1, \dots, u_n) = \Phi_\Sigma\left(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n)\right)$$

$where\ \Phi(x)\ is\ the\ cumulative\ distribution\ function\ of\ a\ \mathcal{N}(0,1)\ normal\ distribution$

$$and\ \Phi_\Sigma\ is\ the\ c.d.f\ of\ a\ multivariate\ \mathcal{N}_n(0_n,\Sigma)\ normal\ distribution.$$

Simulation of a Gaussian Copula:

a. Simulation of $Z = (Z_1,\dots,Z_n) = \begin{pmatrix} Z_1^1 & \cdots & Z_n^1 \\ \vdots & \ddots & \vdots \\ Z_1^{ns} & \cdots & Z_n^{ns} \end{pmatrix} \in \mathbb{R}^{n\times ns}$, of independent standard normal variables ($\forall i = 1,\dots,n\ Z_i \sim \mathcal{N}(0,1)$).
b. Multiplication of Z with the lower triangular matrix of the Cholesky's decomposition of the correlation matrix Σ. We have then $Y = ZC$, if $C\ is\ defined\ such\ that\ CC^t = \Sigma$.
c. The simulated Gaussian copulas are:

$$Y = (Y_1,\dots,Y_n) = \begin{pmatrix} Y_1^1 & \cdots & Y_n^1 \\ \vdots & \ddots & \vdots \\ Y_1^{ns} & \cdots & Y_n^{ns} \end{pmatrix} \in \mathbb{R}^{n\times ns}, with\ \forall i \in [\![1,ns]\!]:$$

$$\left(Y_1^i,\dots,Y_n^i\right) \sim \mathcal{N}_n(0_n, CC^t = \Sigma) \in \mathbb{R}^n$$

There are alternative to the Gaussian copula. It is the Student copula. The particularity of those two copulas is that they can be characterized with a correlation matrix. Both belong to the family of elliptical copulas. The main advantage of the Student copula is the existence of a degree of freedom parameter. It allows us to better study the (right) tail dependencies.

Definition: Student Copula

$$C(u_1,\dots,u_n) = \Phi_{\Sigma,\mathrm{d}}\left(\Phi_d^{-1}(u_1),\dots,\Phi_d^{-1}(u_n)\right)$$

$$where\ \Phi_d(x)\ is\ the\ c.d.f\ of\ a\ Student\ distribution\ with\ d\ degrees\ of\ freedom$$

$$and\ \Phi_{\Sigma,\mathrm{d}}\ is\ the\ c.d.f\ of\ a\ multivariate\ Student\ distribution\ with\ \Sigma\ and\ d\ degrees\ of\ freedom$$

Simulation of a Student Copula:

a. Simulation of $Z = (Z_1,\dots,Z_n) = \begin{pmatrix} Z_1^1 & \cdots & Z_n^1 \\ \vdots & \ddots & \vdots \\ Z_1^{ns} & \cdots & Z_n^{ns} \end{pmatrix} \in \mathbb{R}^{n\times ns}$, of independent standard normal variables ($\forall i = 1,\dots,n\ Z_i \sim \mathcal{N}(0,1)$).

b. Multiplication of Z with the lower triangular matrix of the Cholesky's decomposition of the correlation matrix Σ. We have then $Y = ZC$, if C $is\ defined\ such\ that\ CC^t = \Sigma$.

c. The simulated correlated Gaussian variables are:

$$Y = (Y_1, \dots, Y_n) = \begin{pmatrix} Y_1^1 & \cdots & Y_n^1 \\ \vdots & \ddots & \vdots \\ Y_1^{ns} & \cdots & Y_n^{ns} \end{pmatrix} \in \mathbb{R}^{n \times ns}, with\ \forall i \in [\![1, ns]\!]:$$

$$(Y_1^i, \dots, Y_n^i) \sim \mathcal{N}_n(0_n, CC^t = \Sigma) \in \mathbb{R}^n$$

d. Simulation of ns Chi-squared independent random variables with ν degrees of freedom: $\chi = \begin{pmatrix} \chi_1 \\ \vdots \\ \chi_{ns} \end{pmatrix}$

e. The simulated Student copulas are:

$$T = (T_1, \dots, T_n) = \begin{pmatrix} T_1^1 & \cdots & T_n^1 \\ \vdots & \ddots & \vdots \\ T_1^{ns} & \cdots & T \end{pmatrix} = \begin{pmatrix} Y_1^1 \sqrt{\frac{\nu}{\chi_1}} & \cdots & Y_n^1 \sqrt{\frac{\nu}{\chi_1}} \\ \vdots & \ddots & \vdots \\ Y_1^{ns} \sqrt{\frac{\nu}{\chi_{ns}}} & \cdots & Y_n^{ns} \sqrt{\frac{\nu}{\chi_{ns}}} \end{pmatrix} \in \mathbb{R}^{n \times ns}, with$$

$$(T_1^i, \dots, T_n^i) \sim t(\Sigma, \nu) \in \mathbb{R}^n$$

The next section will describe how to implement a copula approach in the aggregation.

7. Application of the copula theory

In order to aggregate risks we use set of Monte Carlo runs for each risk. In general, we simulate 500k simulations in each run. We take into account the diversification, i.e., the correlation matrices are used during the aggregation. For that we need the correlation matrices to have certain properties, like the positive definiteness[22] of the correlation matrix.

The aggregation is based on a Gaussian Copula method. We use the Cholesky's decomposition of the correlation matrix to correlate Gaussian vectors using the expert's judgment correlation.

Here is the procedure for aggregating n risks $(X_1, \dots, X_n)$

[22] See the section about the Correlation matrix for more details.

1. Simulate for each risk a Monte Carlo run of ns simulations (usually, we take $ns = 500'000$.

$$(X_1, \dots, X_n) = \begin{pmatrix} X_1^1 & \cdots & X_n^1 \\ \vdots & \ddots & \vdots \\ X_1^{ns} & \cdots & X_n^{ns} \end{pmatrix} \in \mathbb{R}^{n \times ns}$$

2. Simulate the Gaussian copulas: $Y = (Y_1, \dots, Y_n) = \begin{pmatrix} Y_1^1 & \cdots & Y_n^1 \\ \vdots & \ddots & \vdots \\ Y_1^{ns} & \cdots & Y_n^{ns} \end{pmatrix}$ (see the section on Copula approach).
3. We apply here the rank method to aggregate the risks, first we create the vector of rank:

$$R = (R_1, \dots, R_n) = \begin{pmatrix} Y_1^{[1]} & \cdots & Y_n^{[1]} \\ \vdots & \ddots & \vdots \\ Y_1^{[ns]} & \cdots & Y_n^{[ns]} \end{pmatrix} \in \mathbb{R}^{n \times ns}, with\ \ Y_j^{[i]}\ the\ rank\ of\ Y_j^i in\ the\ vector\ Y_j$$

For each risk j, we link the rank of Y_j and X_j. We rearrange the simulations in each Monte Carlo run of the risks with the following approach:

$$\left(X_{[1]}, \dots, X_{[n]}\right) = \begin{pmatrix} X_1^{R_1^1} & \cdots & X_n^{R_n^1} \\ \vdots & \ddots & \vdots \\ X_1^{R_1^{ns}} & \cdots & X_n^{R_n^{ns}} \end{pmatrix}^{23} \in \mathbb{R}^{n \times ns}, with\ \ X_j^{[i]}\ the\ rank\ of\ X_j^i in\ the\ vector\ X_j$$

4. We then sum the rearranged values of the Monte Carlo simulations row by row. Each summed row is a simulation of a Monte Carlo run of the aggregated capital charge.

The steps are the same if one want to aggregate n risks $(X_1, \dots, X_n)$ with the Student copulas.

The Positive Definiteness of the Correlation Matrix is important, in fact, AXA's copula approach we need to have Positive Definite matrices in order not to have degenerated multivariate distributions. The section 5 of this chapter describes two algorithms that can be applied to a non-PD matrix to transform it into a PD one.

8. Adequacy of the aggregated capital charge

We also studied the convergence at the aggregated level. After computing all the capital of the standalone scenarios, we computed the aggregated capital charge. Here, we took the 5

[23] We remind that $R_j^i = Y_j^{[i]}$.

standalone scenarios that use a Poisson distribution for the frequency and a lognormal distribution for the severity. To study the convergence of the aggregated capital charge, we fixed the simulations of the five standalone scenario. We want to see the impact of the simulated Gaussian Copulas, by simulating at each time new copulas, which are needed to aggregate the scenarios.

We have simulated 1000 times the aggregated capital charge (obtained with the aggregation of the five scenarios with 500'000 simulations). We can see that the sample does not deviate too much from to the mean.

Minimum	1st Quartile	Median	3rd Quartile	Max	Mean	Std Dev
8 408 544	8 482 106	8 507 955	8 535 001	8 655 857	8 509 919	37 294

Table 18 Descriptive statistics of aggregated Monte Carlo runs

$$\frac{\sigma(STEC)}{\mu(STEC)} = \frac{37\ 294}{8\ 509\ 919} = 0{,}45\%$$

To conclude this section, taking 500'000 simulations in each Monte Carlo run for the scenarios is adequate. The results at the aggregated level are satisfying, we have a good level of precision and the convergence is demonstrated.

Chapter 6: Comparison of the Scenario Based Approach with a Loss Distribution Approach

The aim of this chapter is to present with more details the Loss Distribution Approach. We will use AXA's internal loss data in order to compare the obtained LDA SCR with the Operational Risk SCR obtained with the SBA. We will try to explain these differences. The statistical tools used on this chapter will be presented such as the non-parametric tests (Kolmogorov-Smirnov, Anderson-Darling, etc.). We will base our study on the same scope of time that we used for the correlations study (geographical and between operational risk categories). The methodology behind a classic LDA will first be presented and an alternative to the collecting threshold problem will be introduced as well. Also, an application of the Extreme Value Analysis (EVA) will be presented as well.

1. Statistic tools

The most popular non-parametric tools are the Kolmogorov-Smirnov test (also written K-S test in the literature) and the Anderson-Darling test. The Pearson's chi-squared test will also be presented in this section.

For starter, the empirical distribution function F_n for n iid observation X_i is defined as follow:

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{[-\infty,x]}(X_i)$$

Where $X_i \sim X$ with X being the random variable of the severity distribution.

The first two tests are the "usual tests" for the severity distributions. For the frequency distribution, we will use the last one, the Pearson's chi-squared test. We will take an interval of confidence of 95%.

Another statistical tool used in this chapter is the QQ-plot. This is a graphical tool that evaluates the adequacy of a distribution with respect to the sample of data tested. We compare the quantiles of the sample (or empirical quantile) with the theoretical quantiles (here, computed with the LDA calibration). If the distributions are similar, each empirical quantile will be correctly estimated with the theoretical quantile of the same level. In practice, for fixed parameters we simulate a sample according a distribution and then we reorder all the simulations. We compare then the reordered simulations with the values of the reordered sample. If the two distributions are alike, then the set of data on the graph will form the first bisector ($y = x$). We will then compare the positions of the points on the graph with respect to the first bisector.

a) Kolmogorov-Smirnov

The K-S statistic quantifies a distance between the empirical distribution function of the sample and the theoretical cumulative distribution function of the tested distribution.

<u>K-S statistic:</u>

Let's denote F a cumulative distribution function. Given a sample (F_n being its empirical distribution function) of n iid observations X_i:

$$D_n = \sup_x |F_n(x) - F(x)|$$ [24]

The K-S statistic can be applied to test both the frequency and severity distributions.

b) Anderson-Darling

Let's denote F the cumulative distribution function to be tested and F_n the empirical distribution function of the n-sample of iid observations X_i. The principle is the same as in the KS statistic test:

$$AD = -n - \sum_{i=1}^{n} \frac{2i-1}{n}\left[\ln\left(\Phi\left(X_{[i]}\right)\right) + \ln\left(1 - \ln\left(\Phi\left(X_{[n+1-i]}\right)\right)\right)\right]$$

With

$rank\left(X_{[i]}\right) = i, \forall i \in [\![1, n]\!]$ *and* Φ *the cumulative distribution function to be tested*

This test does not require calibrating any parameters for the distribution to be tested.

c) Pearson's chi-squared test

The Pearson's chi-squared test presented in this section is the test for the "adequacy" of the distribution. It is different from the Pearson's chi-squared test of independence.

The statistic is defined as following:

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = N \sum_{i=1}^{n} \frac{(O_i/N - p_i)^2}{p_i}$$

With

[24] Note that if F_n is drawn from the F then $D_n \to 0$ by the Glivenko-Cantelli theorem.

$$\begin{cases} \chi^2 & Pearson's\ cumulative\ test\ statistic, which\ asymptotically\ approaches\ a\ \chi^2\ distribution \\ O_i & the\ number\ of\ observation\ of\ type\ i \\ N & total\ number\ of\ observations\ in\ the\ sample \\ E_i & expected\ (theoritical) frequency\ of\ type\ i^{25} \\ n & number\ of\ different\ values\ in\ the\ sample \end{cases}$$

This statistic can be used to compute a $p-value$ by comparing the statistic value to a chi-squared distribution. The number of degrees of freedom is $n-p$. With p being the reduction in the degrees of freedom. We have that $p=s+1$ where s is the number of parameters of the distribution to be tested. For a lognormal or Weibull distribution, we have that $p=3$.

2. Classic LDA

This approach would be by far the most used and popular. The principle is fitting statistical distributions with available historic data (either internal or external). More precisely, we compute a frequency and severity distribution with which we recreate a total losses distribution.

$$S=\sum_{i=1}^{N} X_i$$

W here N is the random variable representing the frequency $\forall i \in [\![1,N]\!]$ $X_i \sim X$ where X is the random variable representing the severity and S is the total loss. It is a purely statistical approach, with no room for subjectivity. We retrieve thanks to the internal data the frequency and severity distributions.

The distributions to be tested are calibrated according to the data. The two most common calibration approaches are:

- The method of moments: the parameters of the distribution to be tested are estimated by equalizing the theoretical moments and the empirical moments. The method is based on the Law of large numbers.
- The maximum of likelihood estimation (MLE): this method gives us better results than the method of moments, which is easier to implement. We estimate the parameters of the distribution to be tested by maximizing the likelihood function.

In a LDA, the frequency is calibrated separately from the severity. A Monte Carlo approach can be used to quantify the total loss distribution.

[25] There is a null hypothesis that the fraction of the type *i* in the population is p_i.

a) Frequency distribution

For each category of risk, we will use monthly number of losses. Our scope of time is January 2008 to December 2014, so we have 84 numbers of losses for each category. Three frequency distributions are tested for each category: the Poisson distribution, binomial distribution and negative binomial distribution. For confidentiality purpose, the real figures have been multiplied by a factor $\alpha \in \mathbb{R} \setminus \{0\}$.
In this section, as we use a classic LDA, we will model the frequency of the Category 1 and 2. We have the following results, and by far, the negative binomial distribution is the most convenient and adequate with respect to the Category 1 and 2:

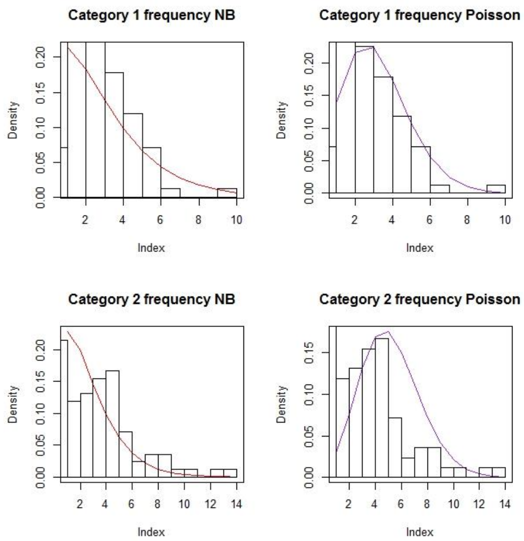


Figure 16 Frequency modelling for Category 1 and 2

We have the below tables of tests' statistics. The Poisson and binomial distributions have similar results, only the Poisson distribution results are presented.

	Poisson (lambda)	Negative Binomial (r,p)	
Parameters	3.106904	2.093012	0.4455548
p-value	0.00800	0.8388632	
Chi-squared	20.69501	3.465527	

Table 19 Category 1 tests' statistic table. Frequency distribution.

	Poisson (lambda)	Negative Binomial (r,p)	
Parameters	5.165713	2.423395	0.318337
p-value	3.35e-08	0.8471343	
Chi-squared	59.05463	6.376026	

Table 20 Category 2 tests' statistic table. Frequency distribution.

For the two categories of operational risk, the negative binomial distribution seems the more appropriate. The table of the statistics is important because we cannot graphically determine the more appropriate distribution. The Poisson distribution and binomial distribution have both very small $p - values$ when fitted with the Pearson's chi-squared test.

b) Severity distribution

For the severity distributions, we use the K-S test and Anderson-Darling test. All the losses (above 0€) were taken into account to model the Category 1 and 2 severity distributions. As in the frequency distribution modelling with a LDA, we multiply our figures by a factor $\alpha \in \mathbb{R} \setminus \{0\}$. The distribution tested here are the following: exponential, lognormal and Weibull.

	Exponential (lambda)	Lognormal (mu, sigma)		Weibull (shape, scale)	
Parameters	4.8387e-05	7.04257742	2.38716316	0.4090138	3'578.75
KS statistic	13.145	0.82218		1.9217	
KS p-value	2.2e-16	0.17		< 2.2e-16	
AD statistic value	Error during the computation in R	2.6081		243.57	
AD p-value	Error during the computation in R	0.33		< 2.2e-16	

Table 21 Category 1 tests' statistic table. Severity distribution.

	Exponential (lambda)	Lognormal (mu, sigma)		Weibull (shape, scale)	
Parameters	6.500020e-05	7.41533267	1.95053179	0.4822521	4'214.520
KS statistic	13.145	0.87579		1.8547	
KS p-value	2.2e-16	0.06		< 2.2e-16	
AD statistic value	19.672	3.5246		479.03	
AD p-value	< 2.2e-16	0.17		< 2.2e-16	

Table 22 Category 2 tests' statistic table. Severity distribution.

The results are satisfying for the lognormal distribution for both categories tested. The exponential distribution and the Weibull distribution don't have a good fit with the Category 1 and 2 data. To confirm graphically these results, we have the following graphs:

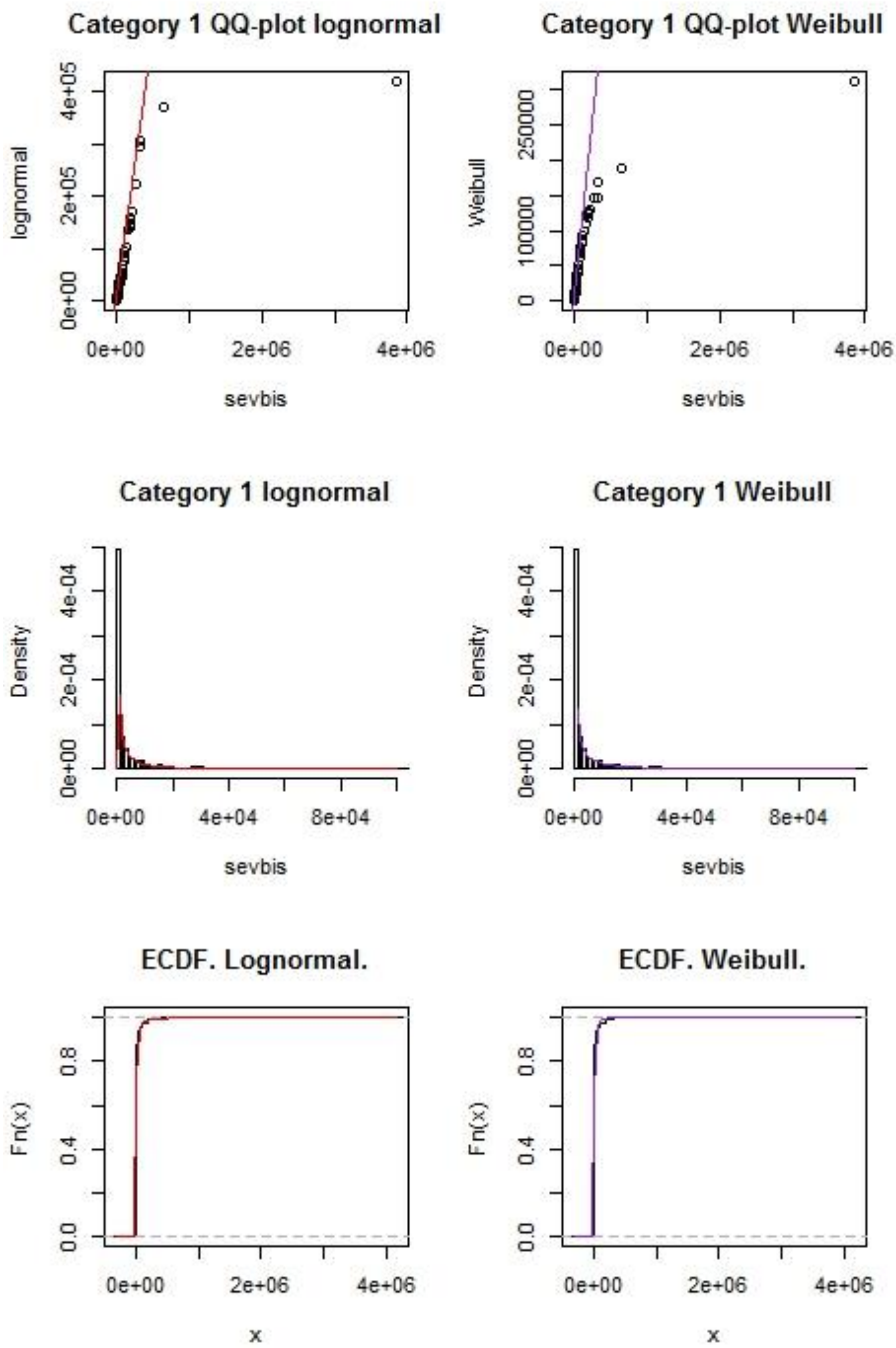


Figure 17 Severity modelling for the Category 1

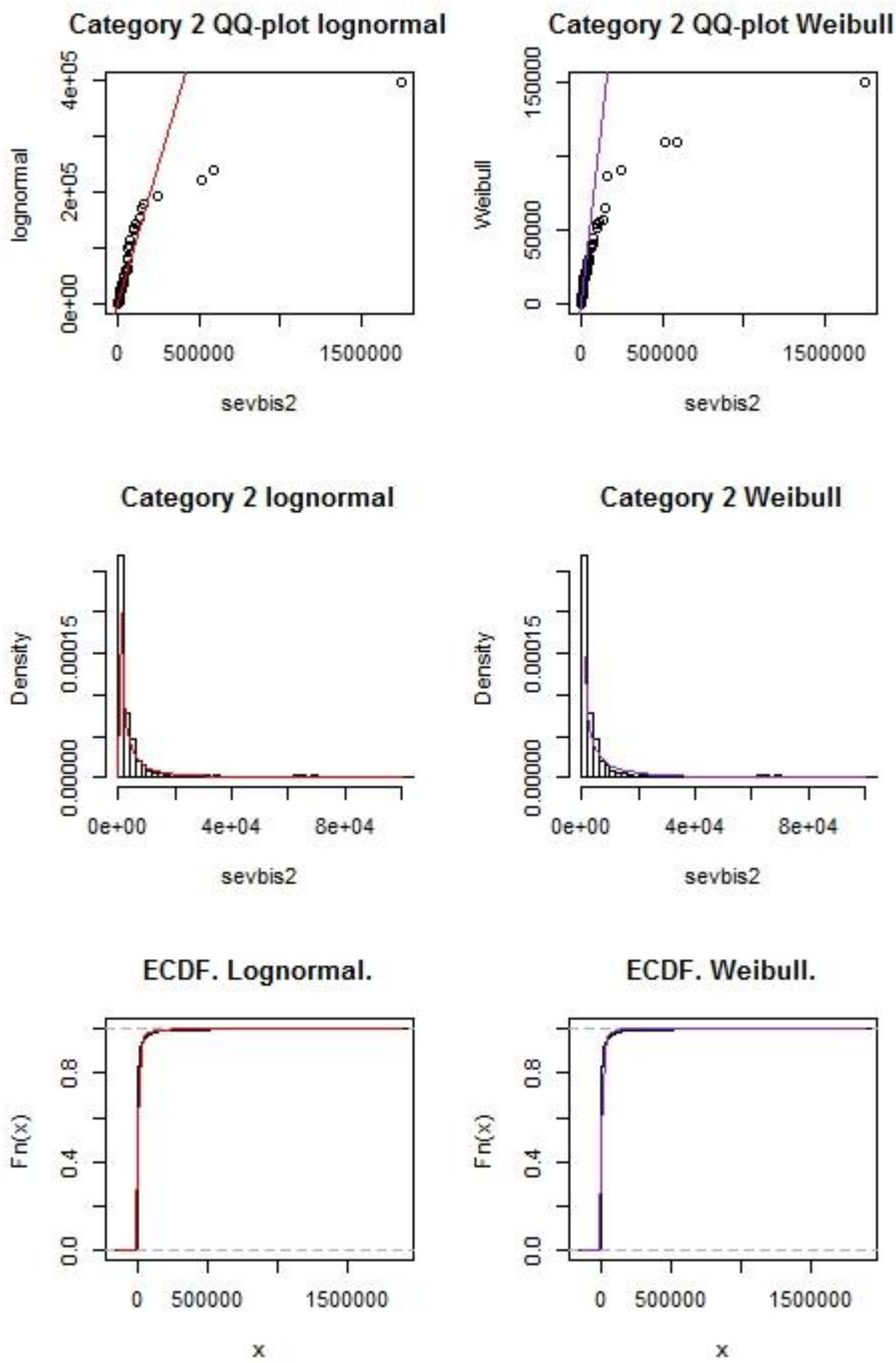


Figure 18 Severity modelling for Category 2

Here for the QQ-plot, the theoretical quantiles are on the y-axis and the sample quantiles on the x-axis. Except for the QQ-plot, we do not see graphically which one of the two distributions is the more adequate. For the Category 1, the QQ-plot gives us good results except on the right tail of the distribution. We are below the bisector, which means that the model underestimates the capital charge for some extreme events. For the Category 2, we are below the bisector around the $quantile(80\%)$, which means we have a lot of "chances" to have a much lower capital charge with the LDA calibration for the Category 2.

3. LDA with a certain threshold

In the Operational Risk industry, all the losses are not always declared. In fact, only losses above a certain threshold are observable in companies' loss database. The threshold is set due the fact that only severity losses above this threshold are collected. In fact, what we were estimating until now are not the cumulative distribution functions but the conditional truncated distribution over a threshold.

If a threshold is applied to the severity distribution X, we have with the previous section approach the parameters calibrated according to the conditional distribution. In fact, we calibrate the parameters of $X|X \geq H$ with H being the threshold. The cumulative distribution function is defined as following:

$$\mathbb{P}(X \leq x|X \geq H) = \frac{\mathbb{P}(H \leq X \leq x)}{\mathbb{P}(X \geq H)} = \frac{F(x) - F(H)}{1 - F(H)}, \forall x > H$$

And $\mathbb{P}(X \leq x|X \geq H) = 0, \forall x \leq H$

The approaches to calibrate the parameters for the conditional distribution are the same. We need also to take into account that a threshold was applied to the severity distribution when modeling the related frequency distribution. In fact, the frequency s underestimated when a threshold is applied. The non-declared losses have to be taken into account in the frequency distribution. We need for that the distribution of the severity X that we estimate with the function "fitdistr" in R. This function allows calibrating a distribution according to a sample of losses above a threshold.

If n is the total number of losses, without any threshold, then we can write $n = n_{\{\leq H\}} + n_{\{>H\}}$ with H being the threshold. To retrieve n we need to first estimate the severity distribution of the losses. For that, we estimate the conditional distribution, and then, we use the function "fitdistr" that implements a MLE approach to calibrate the parameters of the distribution. The frequency distribution with threshold is then adjusted with respect to the severity distribution without threshold.

We will apply the LDA with a threshold to the Category 4 of operational risk. We remind that the factor $\alpha \in \mathbb{R} \setminus \{0\}$ is applied to all the losses here.

In the below table, all the distributions tested are not adequate. All the $p - values$ are close to 0. We can say that the lack of data does not allow us to retrieve the severity distribution with precision. We then plot the QQ-plot, the cumulative distribution function and the density function of the lognormal and Weibull distribution.

	Exponential (lambda)	Lognormal (mu, sigma)		Weibull (shape, scale)	
Parameters	1.209661e-06	7.68537314	2.90650962	0.4049857	8'704.962
KS statistic	9.1731	4.5917		12.9273	
KS p-value	< 2.2e-16	< 2.2e-16		< 2.2e-16	
AD statistic value	Error during the computation in R	186.9		250340	
AD p-value	Error during the computation in R	< 2.2e-16		< 2.2e-16	

Table 23 Category 4 tests' statistic table. Severity distribution.

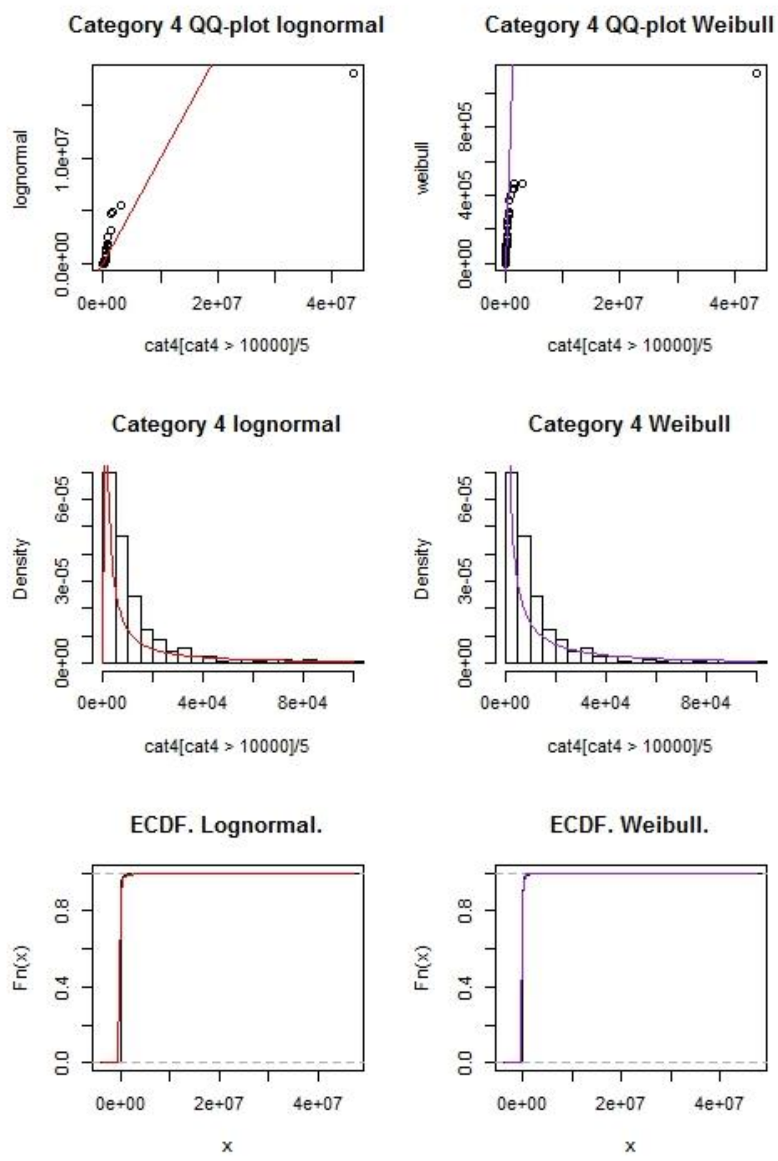


Figure 19 Severity modelling for the Category 4

The lognormal QQ-plot seems to be more conservative at the beginning of the distribution. Nevertheless, we are still less conservative regarding the right tail of the distribution. We then calibrate the frequency distribution according to the severity distribution calibrated.

	Poisson (lambda)	Negative Binomial (r,p)	
Parameters	6.443167	16.53532	0.7133968
p-value	0.6778769	0.8427173	
Chi-squared	7.375899	5.663255	

Table 24 Category 4 tests' statistic table. Frequency distribution.

Here, both the Poisson and negative binomial distribution "fit" the sample. As the $p-value$ of the negative binomial distribution is higher, we would naturally select this distribution. The graph of the Poisson distribution seems more adequate, but since the negative binomial $p-value$ is higher, we select the negative binomial distribution.

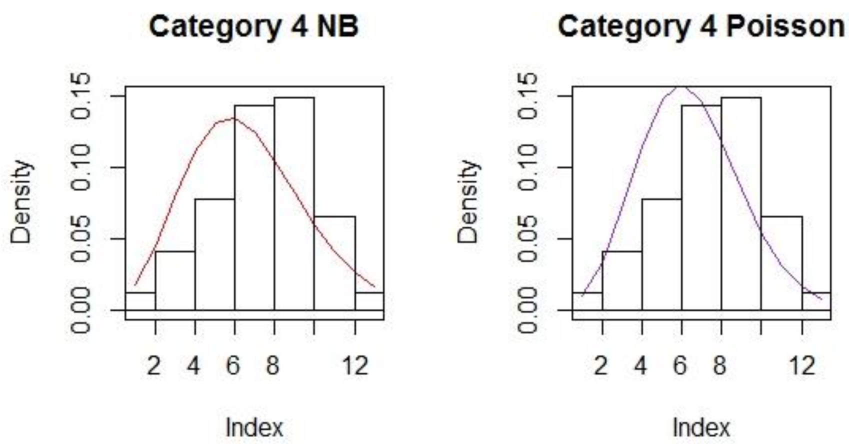


Figure 20 Category 4 Frequency modelling

We see in this part that the lack of coherent data influence greatly the calibration of the frequency severity model for the category 4 of operational risk. Even with the calibration with the threshold taken into account does not give us good results. The next section will now compare the capital charges of the operational risk categories that have coherent data with the two approaches: the SBA and the LDA.

4. Comparison of two approaches results: SBA and LDA

We will compare in this section the results of the capital charge obtained with a LDA and a SBA. Due to some categories that don't have enough "good results" during the calibration with the LDA, we will only compare the categories of operational risk that had both $p-values$ for the frequency and severity distribution superior to 0.05. The operational risk categories selected here are:

- Category 1
- Category 2
- Category 3

The Category 4 to 7 had at least the frequency distributions test that gave $p-values$ inferior to 0.05 or the severity distributions tests that gave $p-values$ inferior to 0.05 (in most of the cases both frequency and severity distributions had $p-values$ inferior to 0.05).

We simulated 1'000 times the $VaR_{99.5\%}$ of the above operational risk categories distribution with the frequency and severity distribution calibrated with the above methodology. We compare after the mean of the 1'000 $VaR_{99.5\%}$ obtained with a LDA with the sum of all the $VaR_{99.5\%}$ scenarios obtained with a SBA. We do not apply any diversification due to the fact that we took all the losses available in the scope of time.

a) Category 1

Let's denote $SCR_{cat\,1}(SBA)$ the capital charge obtained with the sum of all the $VaR_{99.5\%}$ of the category 1 scenarios. Let's denote $SCR_{cat\,1}(LDA)$ the capital charge obtained with a LDA. We have that:

$$\frac{SCR_{cat\,1}(LDA)}{SCR_{cat\,1}(SBA)} = 38.50\%\ and\ \frac{SCR_{cat\,1}(LDA) - SCR_{cat\,1}(SBA)}{SCR_{cat\,1}(SBA)} = -61.50\%$$

The capital charge obtained with a SBA seems more conservative. In fact, many extreme events are taken into account in the scenarios. In the LDA, we do not have any on the internal database these types of extreme loss. Even if $SCR_{cat\,1}(LDA) < SCR_{cat\,1}(SBA)$, we have that the losses related to the Category 1 during the Year 2014 represent only **2.27%** of the $SCR_{cat\,1}(LDA)$ and in average during the period 2008-2014 we have that the annual loss represent **5.67%**.

Year	Percentage of the SCR(LDA) Cat 1
2008	4.34%
2009	17.22%
2010	6.15%
2011	3.27%
2012	4.99%
2013	1.42%
2014	2.27%
Average	**5.67%**

Table 25 Ratio of the total annual loss and the SCR(LDA) of the Category 1 (total annual loss/SCR(LDA))

In fact, in the Category 1's calibration with a SBA, many assumptions are made for extreme events. We need to take them into account, but we also need to put a threshold. All Insurance companies certainly do not need a scenario with a potential extreme event set at \$18bn for this category (potential fine of Volkswagen's operational error). Good risk awareness can avoid these types of event. Also, with the QQ-plot, we know that the right tail of the distribution was underestimated. Thus, the optimal $SCR_{cat\ 1}$ could be between the two capital charges $SCR_{cat\ 1}(LDA)\ and\ SCR_{cat\ 1}(SBA)$.

b) Category 2

Let's denote $SCR_{cat\ 2}(SBA)$ the capital charge obtained with the sum of all the $VaR_{99.5\%}$ of the Category 2 scenarios. Let's denote $SCR_{cat\ 2}(LDA)$ the capital charge obtained with a LDA. We have that:

$$\frac{SCR_{cat\ 2}(LDA)}{SCR_{cat\ 2}(SBA)} = 2.09\%\ and\ \frac{SCR_{cat\ 2}(LDA) - SCR_{cat\ 2}(SBA)}{SCR_{cat\ 2}(SBA)} = -97.91\%$$

In this Category, we see that the lack of data clearly gives us an underestimated capital charge. It is absurd to reduce the capital charge of the Category 2 by 97.91% even if another approach is used. In fact, just in 2014 for the Category 2, the losses represent **<u>14.22%</u>** of the $SCR_{cat\ 2}(LDA)$. From 2008 to 2013, the annual losses are even more important:

Year	Percentage of the SCR(LDA) Cat 2
2008	95.23%
2009	12.50%
2010	32.74%
2011	50.73%
2012	51.89%
2013	19.38%
2014	14.22%
Average	**39.53%**

Table 26 Ratio of the total annual loss and the SCR(LDA) of the Category 2 (total annual loss/SCR(LDA))

The Operational Risk SCR represents a potential loss that can occur once every 200 years. In the above table, we see that in 2008 the total annual loss is almost equal to the capital charge. It is without taking the figures "as if". Another table that can be studied is the quantile that represents the annual loss of the years 2008-2014. The $SCR_{cat\ 2}(LDA)$ represent the quantile at a 99.5%. With the Monte Carlo simulations, we can retrieve the quantile that corresponds to the annual loss of the considered year. According to the LDA calibration, we have the following table:

Year	Annual loss corresponding quantile	Annual loss that occurs once every
2008	99,44%	179,533214
2009	72,78%	3,6734454
2010	93,69%	15,8443452
2011	97,52%	40,3975115
2012	97,65%	42,5604358
2013	84,35%	6,39157335
2014	76,44%	4,24477044

Table 27 Annual losses corresponding quantiles and occurrences Category 2

In 2008, the total annual loss is almost corresponding to the $SCR_{cat\ 2}(LDA)$. In the above table, we see that according to the LDA, the total losses of 2008 should have happened only once every 180 years. In 2011-2012 we have for two consecutive years a total annual loss that should have happened once every 40 years. We conclude that the set of data used is insufficient to do a proper LDA for the Category 2.

For the Category 2, even with "good results" in the calibration part of the LDA, we do not have a good quantification of this category of operational risk. In this case, the use of a SBA is preconized.

c) Category 3

The Category 3 of operational risk has like the Category 1 & 2 "good results" on the calibration with a LDA. We multiplied our figures by a factor $\alpha \in \mathbb{R} \setminus \{0\}$ for the tables.

	Poisson (lambda)	Negative Binomial (r,p)	
Parameters	1.190184	0.499813	0.3413804
p-value	0.004720519	0.8248528	
Chi-squared	14.99071	0.9023825	

Table 28 Category 3 tests' statistic table. Frequency distribution.

The negative binomial distribution has a higher $p-value$, thus is more adequate. Also, the $p-value$ of the Poisson distribution is below 0.05 which is our threshold for the 95% confidence interval.

	Exponential (lambda)	Lognormal (mu, sigma)		Weibull (shape, scale)	
Parameters	5.532880e-05	9.17835542	1.06468483	0.8838492	16'114.57
KS statistic	1.9497	1.135		1.8873	
KS p-value	< 2.2e-16	0.03		< 2.2e-16	
AD statistic value	5.7836	3.5376		5.2765	
AD p-value	0.08	0.12		0.06	

Table 29 Category 3 tests' statistic table. Severity distribution.

Here we have all the AD $p-values$ that are above 0.05. We will then use the negative binomial and the lognormal distribution to compute the operational risk capital charge of the Category 3.

Let's denote $SCR_{cat\ 3}(SBA)$ the capital charge obtained with the sum of all the $VaR_{99.5\%}$ of the Category 3 scenarios. Let's denote $SCR_{cat\ 3}(LDA)$ the capital charge obtained with a LDA. We have that:

$$\frac{SCR_{cat\ 3}(LDA)}{SCR_{cat\ 3}(SBA)} = 0.40\%\ and\ \frac{SCR_{cat\ 3}(LDA) - SCR_{cat\ 3}(SBA)}{SCR_{cat\ 3}(SBA)} = -99.60\%$$

We have similar results with the Category 2. Here, the Operational Risk SCR of the category is way too underestimated. We will compare the total annual losses of the year 2008-2014 and compare them to the $SCR_{cat\ 3}(LDA)$:

Year	Percentage of the SCR(LDA) Cat 3
2008	86,82%
2009	89,07%
2010	108,92%
2011	116,47%
2012	74,96%
2013	31,65%
2014	13,61%
Average	**74,50%**

Table 30 Ratio of the total annual loss and the SCR(LDA) of the Category 3 (total annual loss/SCR(LDA))

In the past years, we had two times in 2010 and 2011 an annual loss that is superior to the $SCR_{cat\ 3}(LDA)$ computed with the calibrated distribution via the LDA. Another table that can be studied (like with the Category 2) is the quantile that represents the annual loss of the years 2008-2014. The $SCR_{cat\ 3}(LDA)$ represent the quantile at a 99.5%. With the Monte Carlo simulations, we can retrieve the quantile that corresponds to the annual loss of the considered year.

Year	Annual loss corresponding quantile	Annual loss that occurs once every
2008	99,02%	102,0824826
2009	99,13%	115,2870648
2010	99,71%	342,7004798
2011	99,83%	584,1121495
2012	98,09%	52,36150382
2013	77,68%	4,479403702
2014	42,89%	1,751086549

Table 31 Annual losses corresponding quantiles and occurrences Category 3

In 2010 and 2011, according to the LDA calibration, the total annual losses should have occurred once every 342 and 584 years. As with the Category 2, we conclude that the LDA is not adequate to calibrate this category of risk. The result here was not totally unexpected. Only the Anderson-Darling's test $p-value$ was above 0.05, and not the KS's test $p-value$. We have higher $p-values$ with the two first categories tested.

5. Conclusion

To conclude this chapter, the LDA can be applied when the internal data loss collection is coherent and mature, but it is not sufficient. Even with "good results" with the statistical tests, we have a SCR of lower level, especially for the category 2 and 3. We clearly see the effects of the scarcity and heterogeneity of the data. The collection of data for some categories is not mature and adapted to use a LDA. In this situation, the SBA is particularly efficient, because we can take into account events that are likely to happen once every 200 years.

For the Category 1 we have a calibration that seems to be adequate, it is the Category where the statistical tests had the best results in the LDA calibration. But we have still a low level capital charge. Even in the Category 1, we do not have enough historical data on extreme events. It is shown on the QQ-plot. The right tail of the distribution is underestimated, and that corresponds to the low-level high frequency risks.

An alternative to the LDA for a good calibration would be to combine this approach with the SBA. For example, we can calibrate the frequencies distribution with the LDA, and the severity distribution with a SBA. In fact, we have better results on the frequency calibration than the severity calibration. We can also think of including an external database. It is the subject of DAHEN H. and DIONNE G. paper[26] on scaling methods to use an external database. In fact, depending on the size of the added database, a scaling is necessary to adapt the database to the company risk's profile. In fact, scaling external data can enable to get information on the right tail of a severity distribution that is not included in a company's internal loss database. As an example of external database, the ORX[27] Global Loss Database contains more than 400'000 losses of operational risk and is currently working on establishing the ORX Global Insurance Service. Its total value of losses is around €200bn and can be of great help in the calibration of an internal model with a LDA.

[26] DAHEN H. and GIONNE G., (2008) Scaling Models for the Severity and Frequency of External Operationnal Loss Data

[27] Operational Risk eXchange association

Conclusion

In this thesis, we have first defined what is the Operational Risk and its framework. The two way of quantifying its operational risk have been introduced (standard formula and internal model). The three main approaches for an internal model are the Bayesian approach, the LDA and the SBA, which is AXA's internal model.

The steps for quantifying operational risks with a SBA have been presented from the risk assessment to the quantification of an operational risk. The most used distributions have been presented alongside with the quantile-matching calibration for the severity distributions. Many points were studied in this thesis. AXA's approach for computing the total loss distribution, the Monte Carlo approach, to compute the total loss distribution has been presented. The convergence of this approach is presented in this study to demonstrate the stability and the convergence of the computed capital charges. To avoid having a null $VaR_{99.5\%}$, a study on a frequency threshold is introduced in this thesis. Operational risks are risks with low frequency high severity. The Solvency II directives consider only events that are expected to occur once every two hundred years, but then, an open subject can be introduced: how to we take into account these risks? As said in the risk measures section, the $VaR_{99.5\%}$ is not subadditive. Thus, increasing the frequency can lead to an overestimation of the capital charge, less adequate to the risk profile of AXA. In the aggregation approach section, the geographical correlations and the correlations between operational risk categories have been computed with the internal data to demonstrate we have a low level of correlation. The correlation matrices have received a special focus since they are important in the aggregation process. For that, the needed properties as well as two algorithms that transform non-PD matrices to PD ones have been presented. Then the copula theory to better catch the dependencies between the right tails of the severity distribution has been applied for the aggregation process.

The last chapter has shown us the concrete lack of data and its heterogeneity. Only three categories among the seven operational risk categories have "good results" with the statistical tests done. Nevertheless, the computed capital charges were not at the same level when computed with a LDA. The alternative of combining two or more approaches for the internal model can be adopted. This alternative has been applied and explained in Gamonet's actuarial thesis with a combination of the LDA for the frequency distribution, and a Bayesian network for the severity distribution. At last, another alternative could be the use of an external database with the appropriate scaling methods.

Bibliography

ACPR (2011) – Solvabilité 2 : principaux enseignements de la cinquième étude quantitative d'inpmact (QIS5), March 2001

ARTZNER P., DELBAEN F., EBER J.M, HEATH D. (2009) Coherent measures of risk, Mathematical Finance, 9, No. 3, pp203-228

BASEL COMMITTEE ON BANKING SUPERVISION (2001), Supervisory Guidelines for the Advance Measurement Approaches, June 2011

CHERANCE F. (2013) Operational Risk: An SBA model validation, ISUP, Paris

DAHEN H. and GIONNE G., (2008) Scaling Models for the Severity and Frequency of External Operationnal Loss Data

EIOPA (2011) Report on QIS5 for Solvency II, March 2011

FASSIER J. (2014) Quantification of Operational Risks using a Scenario Based Approach, Université Paris-Dauphine, Paris

GAMONET J. (2010) Modélisation du risque opérationnel dans l'assurance, Centre des Etudes Actuarielles, Paris

HIGHAM N. (2002), Computing the Nearest Correlation Matrix – A problem from Finance, The University of Manchester

KENDALL M., DICKINSON GIBBONS J., (1990), Rank Correlation Methods, Oxford University Press, 5th edition

REBONATO R. JÄCKEL P. (1999) The most general methodology to create a valid correlations matrix for risk management and option pricing purposes

RENAUDIN A. (2012) Modèle de capital économique pour le risque opérationnel bancaire: estimation, diversification, ISFA

FRACHOT A., RONCALLI T., SALOMON E., (2004), The Correlation Problem in Operational Risk, Groupe de Recherche Opérationnelle du Crédit Agricole, France, 2004

R, https://stat.ethz.ch/R-manual/

TAHANI A. MATURI (2010) A Comparison of Correlation Coefficients via a Three-Step Bootstrap Approach, Journal of the Mathematics research Vol. 2, No 2, 2010.

WIKIPEDIA, https://en.wikipedia.org

List of Tables

List of Figures