OPERATIONAL RISK MODELING UNDER THE LOSS DISTRIBUTION APPROACH

A Thesis

presented to

the Faculty of Natural and Applied Sciences

at Notre Dame University-Louaize

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Actuarial Science

by

HRAIR RAZMIG DANAGEUZIAN

JULY 2021

Operational Risk Modeling under the Loss Distribution Approach

By

Hrair Razmig Danageuzian

July 2021

Department of Mathematics and Statistics
Faculty of Natural and Applied Sciences
Notre Dame University – Louaize
Zouk Mosbeh, Lebanon

The undersigned certify that they have examined and approved this thesis in partial fulfillment of the requirements for the degree of Master of Science in Actuarial Science:

Dr. Re-Mi Hage, **Thesis Advisor**
Signature: Re-mi

Dr. Roger Nakad, **Reader and Chair**
Signature: Roger Nakad

This thesis is accepted by the Faculty of Natural and Applied Sciences.

Dr. George Eid, **Dean**
Signature: George H. Eid

ACKNOWLEDGEMENT

I would like to express gratitude to Chedid Re for granting me a scholarship to pursue a Master's in Actuarial Science at NDU. Their generous fund allowed me to join the program and work on this research. This scholarship has allowed me to develop my academic and research background, and I am confident this degree will be a distinctive asset for my future endeavors.

I would like to thank my thesis advisor Dr. Re-Mi Hage, for giving me the opportunity to enter an interesting new field. Her continuous guidance, substantial support, and considerable commitment were the bedrock of this research. Many thanks are also addressed to Dr. George Eid, Dean of the Faculty of Natural & Applied Sciences, and Dr. Roger Nakad, Chairperson of the Department of Mathematics & Statistics and committee member.

In addition, I would like to address special thanks to all the professors at the department who taught me throughout my time at NDU, namely Dr. John Haddad, Dr. Ramez Maalouf, Mrs. Claudia Bou Nassif, and my advisor Dr. Re-Mi Hage. I had my first academic encounter with Dr. Maalouf who taught me a course on Financial Economics for Actuaries. The course was really interesting and it is a great loss not to have him with us. I certainly looked forward for his classes.

ABSTRACT

The term operational risk became widespread in the late 1990s when central bank representatives of twelve countries formed a working committee; the Basel Committee on Banking Supervision (BCBS). The BCBS defines operational risk as the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events. This research aims to model operational risk data using the Loss Distribution Approach under BCBS requirements.

Simulated data was used consisting of 3,192 operational loss events between the years 2009 and 2018. The implementation of the LDA was conducted using R programming language; R studio 4.0.3. Due to the low count of loss events, the LDA could not be implemented at business line-risk category levels. Rather, it was implemented per business line and a second time per risk category. The capital requirement was determined for each case. Loss frequency and severity distributions were modeled, the aggregate loss distribution was determined through convolution, and finally the overall distribution was obtained through a copula function. Capital requirements were calculated for each year as the difference between the 99.9% VaR and the Expected Loss (EL).

Significant differences were identified between the yearly capital requirements obtained for each of the two cases. Since operational risk data encompasses high-frequency low-severity and low-frequency high-severity events, the variations of gross loss amounts within business lines and risk categories have a huge impact on the capital requirement. As per Basel requirements, internally generated operational risk measures used for regulatory capital purposes must be based on a minimum five-year observation period of internal loss data. Therefore, the total 10 year period was considered and a weighted average of the capital charge was calculated. Both cases yielded rather close capital charges. The business line method recorded a lower capital charge by around 15%. Ultimately, and to diminish the impact of operational risk, the larger capital charge of 8,738,614$ is recommended for the next year. The impact of the research findings is correlated towards a better understanding of the composition and distribution of operational risk data over risk classes and the corresponding operational risk capital requirements.

Keywords: Operational Risk, Basel, Loss Distribution Approach, Capital Charge, Convolution, Monte Carlo Simulation, Copula, Value at Risk, R Studio.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

BCBS	Basel Committee on Banking Supervision
BIA	Basic Indicator Approach
SA	Standardized Approach
ASA	Alternative Standardized Approach
AMA	Advanced Measurement Approaches
NSA	New Standardized Approach
VaR	Value at Risk
EVT	Extreme Value Theory
LD	Lognormal Distribution
PD	Poisson Distribution
PRD	Pareto Distribution
GPD	Generalized Pareto Distribution
WD	Weibull Distribution
NBD	Negative Binomial Distribution
LDA	Loss Distribution Approach
IMA	Internal Measurement Approach
SCA	Scorecard Approach
SBA	Scenario Based Approaches
POT	Peak over Threshold
MLE	Maximum Likelihood Estimation
LSE	Least Square Estimation

Chapter One: Introduction

1.1 Background

Operational risk is the risk which outlines the uncertainties and the threats that an organization experiences as it carries out its daily operations and business activities. Power, (2005) considers that businesses, banks in particular, have been aware of uncertainties arising from fraud, business disruption, defective information technology and infrastructure, and legal liability for many years. The assortment of such risks under operational risks has created a separate status which requires managerial and regulatory support. This formation of operational risks has connected good governance to risk management. Even though the term operational risk existed in the early 1990s, it only became widespread in the late 1990s with the development of Basel II proposal when the central bank representatives of twelve countries formed a working committee which was the Basel Committee on Banking Supervision. In 1994, the Basel Committee had already recognized the importance of risks related to business operations mainly as deficiencies in information systems or internal controls. In the banking industry, operational risk started as a residual category, as risks left behind from market and credits risks. Basel II reforms have successfully institutionalized operational risk as a category of regulatory knowledge production. Furthermore, Basel II mirrors an overall climate of regulatory attention to organizational control systems and cultures of control. However, definitional issues, data collection, and limits of quantification were the three main controversies of Basel II reforms to banking supervision.

Balthazar, (2006) highlights the existence of a need in the 1980s for an international regulation to generate a more secure system in financial institutions. It was after the numerous banking crises of the 1980s that capital requirements were imposed as an international benchmark in banks. An overview on the history of banking regulation and bank failures worldwide from the mid-1800s and up until the late 1990s has showed that some common elements were recurrent in each and every banking crisis. These elements include deregulation phases, the entry of new competitors, asset price booms, and tighter monetary policies. History has proved that banks can actually go bankrupt and that there exists an illusory feeling of safety regarding the financial systems of developed countries. Thus, a regulatory framework is vital.

Fritz-Morgenthal, (2015) discusses the evolutionary steps towards operational risk management. Denial is the first step expressed in the form of "there is no such thing as operational risk". In the 1980s, banks considered that they have credit, market, and liquidity risk only. In the 1990s, and as a second step, ignorance emerged. Banks and financial institutions would not acknowledge that they were subject to operational risk. Before the 2000s, originated the third step, that of zero tolerance. They did not accept operational risk in their institutions. With the Basel II framework in the early 2000s, the fourth step of operational risk management began. That is, after banks and financial institutions acknowledged operational risk, they started collecting events and classifying operational risks. In step five, they started measuring operational risk using internal and external approaches. Thus, the management of operational risk became possible. Around the beginning of 2010, and with multibillion losses from operational risks, banks and financial institutions entered the "Wake up" step or phase. They discovered that their approaches in quantifying and modeling operational risk did not actually describe their risk profile. Finally, banks and financial institutions by 2015 understood the importance and complexity of operational risk as an important component of the institution's total risk. They entered the "New Normal" step and addressed how to improve the measurement and management of operational risks.

1.2 Operational Risk Overview

The Basel Committee on Banking Supervision in a revised framework of Basel II (2006), provides a comprehensive definition of operational risk. "Operational risk is defined as the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events. This definition includes legal risk but excludes strategic and reputational risk". Legal risk includes, but is not limited to, exposure to fines, penalties, or punitive damages resulting from supervisory actions, as well as private settlements.

Chaudhuri and Gosh, (2016) consider operational risk as a key risk component for banks and financial institutions. Operational risk is estimated to constitute between 15 and 25 % of total risks and thus requires consideration. In the 1970s and 1980s, derivatives were used to hedge market risk and credit derivatives were used to hedge credit risk. Even the first recommendations of the Basel Committee were not concerned with operational risk, considering that hedging the market and credit risks inevitably covers operational risks. Financial institutions, banks, and insurance companies have experienced more than 100 operational loss events suffering hundreds of millions of dollars in the 1990s and 2000s. Some examples include Allfirst Financial with $691 million rogue trading loss, Household Finance with $484 million settlement due to misleading sales practices, and the Bank of New York with $140 million due to the 9/11 attack. Such issues must be handled with a uniform set of rules. The BCBS provides regulatory frameworks and acts as a forum for regular cooperation on banking supervisory matters. The objective of the committee is to improve the understanding of key supervisory issues and the quality of banking supervision.

1.3 Research Aim and Objectives

This research aims to model operational risk data using the Loss Distribution Approach under the Basel Committee on Banking Supervision requirements. The research aim is achieved through the following objectives:

1. Conduct a comprehensive literature review on operational risk, Basel capital accords, and modeling methods for operational risk losses.
2. Identify Basel requirements and measurement methodologies for calculating operational risk capital.
3. Implement the LDA by modeling frequency and severity distributions, and conducting the convolution and copula methods.
4. Estimate operational risk capital requirements.

1.4 Scope of Work

The scope of work is primarily determined by the operational risk loss data. Even though there are many operational risk databases, the data is not accessible to the public. Banks and other financial institutions join these databases by sharing their own data and thus having access to the community data. Therefore, in this research, simulated data was used which restricted the scope of work:

1. The dataset used included the date of observation, loss amount, and the business line/risk category the loss belonged to. This created numerous limitations in the potential selection of measurement methodologies under Basel requirements.
2. The Basic Indicator Approach and Standardized Approach were not implemented because the data used lacked the annual gross income.
3. The Loss Distribution Approach was implemented with certain setbacks due to the limited number of operational risk loss events in numerous risk classes.
4. Since the frequency of losses was not accessible, therefore modeling frequency distribution was compromised and an alternative method was integrated to produce the required results.

1.5 Research Significance and Contribution

The significance of this research lies in providing a quantitative understanding of operational risk data in banks or financial institutions. This is achieved through the implementation of the LDA consistently with Basel requirements.

The contribution of this research is the estimation of operational risk capital by breaking down the data to all business line and risk category classes. The impact of the research findings is correlated towards a better understanding of the composition and distribution of operational risk data over risk classes and the corresponding operational risk capital requirements.

Numerical results are produced which can be used for future research along with the required codes which could be used as tools to further examine other characteristics of operational risk.

Chapter Two: Literature Review

2.1 Basel Committee on Banking Supervision

2.1.1 The Basel I Capital Accord

The central bank representatives of twelve countries formed a working committee which was the Basel Committee on Banking Supervision. The committee provided recommendations which were first published in 1988. The committee aimed at defining capital requirements based on a bank's balance sheet position. Their initiative held two main objectives which were "to strengthen the soundness and stability of the international banking system, and to diminish existing sources of competitive inequality among international banks".

The Basel I framework was designed in a way to define and impose a minimum capital level on internationally active banks. Countries had the option to implement stronger requirements and to adopt the framework on national banks as well. The first step of the Basel I framework was determination of capital, that is, what is considered as capital (two classes Tier 1 and Tier 2). Next in the framework was the definition of a number of factors that would weigh the balance sheet amounts to reflect their risk levels (five broad categories). The committee, in a final step, defined weighing schemes for the off-balance sheet items which were divided into engagements similar to unfunded credits and derivative instruments. Increased competition and internationalization of the banking industry emphasized the need for market risk capital rules which conveyed the 1996 Market Risk Amendment to the 1988 Basel I framework. A new class of capital (Tier 3) was introduced to support market risk recognizing short-term subordinated debts as capital instruments. The Basel I Accord provided two approaches for the calculation of the required capital. The first is the Standardized Approach in which the capital requirements for interest rates and equity positions are designed to cover only specific risks which are defined as "movements in market value of the individual security owing to factors related to the individual issuer" and general risks which are "the risks of loss arising from changes in market interest rates, or from general market movements in the case of equities". A risk-weight by function of type and maturity is assigned to interest rate sensitive instruments for specific risks. Another capital requirement is estimated by categorizing securities into stacks based on maturity and integrating some recognition of long and short positions in the same currency. The second approach is the Internalized Models Approach which bases the capital requirement on the Value at Risk models which are the bank's proprietary internal models. A pricing model is adopted to value each position where the underlying risk parameters are simulated, the generated outcomes of the risk drivers are injected back in, and all positions are reexamined. Thousands of simulations are done to produce risk metrics which simulate a whole distribution of the potential future values.

The main accomplishment of the Basel I Accord was the establishment of a worldwide benchmark of banking regulations. The Accord imposed a uniform set of rules on the required capital levels of international banks conducting the same business in many different countries. The Accord has created a safer banking sector as the capital ratios of most banks increased in the 1990s, yet it included several regulatory weaknesses. Basel I Capital Accord became less efficient as banks adopted the concept of Capital Arbitrage. Making an arbitrage between regulatory and economic capital to align them more closely allowed banks to correct the weakness of regulatory constraints. Other weaknesses of the Accord include lack of risk sensitivity, limited recognition of collateral, incomplete coverage of risk sources, and no recognition of diversification.

2.1.2 The Basel II Capital Accord

The final proposal of Basel II was published in 2004. It held three main objectives which were "to increase the quality and the stability of the international banking system, to create and maintain a level playing field for internationally active banks, and to promote the adoption of more stringent practices in the risk management field". To meet its objectives, the Accord was developed on the basis of three pillars.

Pillar 1 – Solvency Ratio

Capital is still considered as the main safeguard against losses, however the way assets are weighted has been developed. A standard simplified credit risk model is used to derive the Basel II values which would align the capital requirements to internal economic capital estimates of banks through internal models. In an attempt to yield a more systematic collateral management practice, explicit capital requirements by function of risk levels are introduced. Furthermore, a new requirement is added for operational risk with an explicit capital requirement related to possible losses arising from errors in processes, internal frauds, and information technology problems.

Pillar 2 – Supervisory Review and Internal Assessment

Banks must evaluate their capital requirements in line with the regulatory framework within their risk profile by devising internal systems and models. Also, banks are required to assimilate risks not covered in the Basel II Accord such as risks related to reputation or strategy. Banks are expected to operate with a capital level higher than 8 % (pillar 1 requirement) to accommodate for other sources of risk. Banks and regulators must cooperate on the evaluation of internal models, and the latter can take actions if they consider that capital requirements are not sufficiently met.

Pillar 3 – Market Discipline

Under pillar 3, banks are required to build and periodically (twice a year) publish comprehensive reports on their internal risk management systems. This will allow the market to place additional pressure on banks to advance their risk management practices (Balthazar, 2006).

The BCBS in a revised framework of Basel II (2006), provides three measurement methodologies for calculating operational risk capital which are the Basic Indicator Approach, the Standardized Approach, and the Advanced Measurement Approaches.

The capital requirements calculated under the three aforementioned approaches are multiplied by 12.5 to determine the risk weighted assets. The value 12.5 is the reciprocal of the minimum capital ratio of 8%.

I. Measurement Methodologies

i. Basic Indicator Approach

This approach imposes on banks to hold capital for operational risk "equal to the average over the previous three years of a fixed percentage (α) of positive annual gross income". The approach requires to exclude annual gross incomes for any year if these figures are negative or zero, and necessitates supervisors to study fitting actions under Pillar 2.

$$K_{BIA} = \frac{\Sigma\,(GI_{1\ldots n} \times \alpha)}{n}$$

Where

- "K_{BIA} = the capital charge under the Basic Indicator Approach".
- "GI = annual gross income, where positive, over the previous three years".
- "n = the number of the previous three years for which gross income is positive".
- "α = 15%, which is set by the Committee, relating the industry wide level of required capital to the industry wide level of the indicator".

The gross income, as defined by national supervisors and national accounting standards, is "the net interest income plus net non-interest income". Thus, this measure should:

- Be gross of any provisions (e.g. for unpaid interest);
- Be gross of operating expenses, including fees paid to outsourcing service providers;
- Exclude realized profits/losses from the sale of securities in the banking book; and
- Exclude extraordinary or irregular items as well as income derived from insurance.

The fees which banks receive for outsourcing services are included in gross income. However, securities classified as "held to maturity" and "available for sale" are excluded from gross income.

ii. Standardized Approach

In this approach, bank activities are divided into eight business lines as highlighted below. Gross income serves as a substitute of the likely scale of operational risk exposure within each of the eight business lines. Each business line is assigned a factor (β) which is a "proxy for the industry-wide relationship between the operational risk loss experience for a given business line and the aggregate level of gross income for that business line". Thus, each business line has its own capital charge which is equal to its corresponding gross income multiplied by β. Table 1 shows a detailed mapping of business lines, presenting the sub-level and activity groups of each.

Table 1: Mapping of Business Lines

Level 1	Level 2	Activity Groups
Corporate Finance	Corporate Finance	Mergers and acquisitions, underwriting, privatizations, securitization, research, debt (government, high yield), equity, syndications, IPO, secondary private placements
	Municipal/Government Finance	
	Merchant Banking	
	Advisory Services	
Trading & Sales	Sales	Fixed income, equity, foreign exchanges, commodities, credit, funding, own position securities, lending and repos, brokerage, debt, prime brokerage
	Market Making	
	Proprietary Positions	
	Treasury	
Retail Banking	Retail Banking	Retail lending and deposits, banking services, trust and estates
	Private Banking	Private lending and deposits, banking services, trust and estates, investment advice
	Card Services	Merchant/commercial/corporate cards, private labels and retail
Commercial Banking	Commercial Banking	Project finance, real estate, export finance, trade finance, factoring, leasing, lending, guarantees, bills of exchange
Payment & Settlement	External Clients	Payments and collections, funds transfer, clearing and settlement
Agency Services	Custody	Escrow, depository receipts, securities lending (customers) corporate actions
	Corporate Agency	Issuer and paying agents
	Corporate Trust	
Asset Management	Discretionary Fund Management	Pooled, segregated, retail, institutional, closed, open, private equity
	Non-Discretionary Fund Management	Pooled, segregated, retail, institutional, closed, open
Retail Brokerage	Retail Brokerage	Execution and full service

The total capital charge is calculated as "the three-year average of the simple summation of the regulatory capital charges across each of the business lines in each year. In any given year, negative capital charges (resulting from negative gross income) in any business line may offset positive capital charges in other business lines without limit. However, when the aggregate capital charge across all business lines within a given year is negative, then the input to the numerator for that year will be zero".

$$K_{SA} = \frac{\sum_{\text{years } 1\text{-}3} \max\left(\sum (GI_{1\text{-}8} \times \beta_{1\text{-}8}), 0\right)}{3}$$

Where

- "K_{SA} = the capital charge under the Standardized Approach".
- "$GI_{1\text{-}8}$ = annual gross income in a given year, as defined above in the Basic Indicator Approach, for each of the eight business lines".
- "$\beta_{1\text{-}8}$ = a fixed percentage, set by the Committee, relating the level of required capital to the level of the gross income for each of the eight business lines"; presented in Table 2.

Table 2: Business Lines and their respective Beta Factors

Business Lines	Beta Factors
Corporate Finance (β_1)	18 %
Trading & Sales (β_2)	18 %
Retail Banking (β_3)	12 %
Commercial Banking (β_4)	15 %
Payment & Settlement (β_5)	18 %
Agency Services (β_6)	15 %
Asset Management (β_7)	12 %
Retail Brokerage (β_8)	12 %

Gross income of business lines must be developed based on a bank's specific policies and criteria which must be reviewed for new or changing business activities. Internationally active banks using the Standardized Approach must meet additional criteria set out by the committee. The principles for business line mapping are also identified by the committee.

Alternative Standardized Approach

A bank may be allowed to use the ASA as long as it provides an improved basis for calculating operational risk capital. Under this approach, the operational risk capital methodology is the same as that of the SA with the exception of two business lines which are retail banking and commercial banking. For these two business lines, gross income is replaced by the product of loans or advances and a factor (m). Thus, the operational risk capital charge for retail and (commercial banking) is represented as:

$$K_{RB} = \beta_{RB} \times m \times LA_{RB}$$

Where

- "K_{RB} is the capital charge for the retail banking business line".
- "β_{RB} is the beta for the retail banking business line".
- "LA_{RB} is total outstanding retail loans and advances (non-risk weighted and gross of provisions), averaged over the past three years".
- m is 0.035
- Loans and advances in retail banking consist of the total drawn amounts in the following credit portfolios: retail, purchased retail receivables, and SMEs treated as retail.
- Loans and advances in commercial banking consist of the drawn amounts in the following credit portfolios: corporate, sovereign, bank, specialized lending, purchased corporate receivables, and SMEs treated as corporate.

iii. Advanced Measurement Approaches

The capital requirement under this approach is equal to the risk measure produced by the internal operational risk measurement system of a bank. This approach follows quantitative and qualitative criteria and requires supervisory approval. The qualitative standards include the following:

1. The bank must have an independent operational risk management function. Its main responsibility is designing and implementing operational risk measurement methodologies, and developing strategies to identify, measure, monitor and control operational risk.
2. An integration of the day-to-day risk management processes of the bank with the internal operational risk measurement system including methods for distributing capital to the main business lines.

3. Reports of operational risk exposures and loss experiences must be conveyed regularly to the bank's management which must take appropriate actions accordingly.
4. The bank must ensure compliance with policies, controls, and procedures related to the operational risk measurement system.
5. Regular reviews of the operational risk management processes and measurement systems must be conducted by auditors to ensure that data flows and processes are easily accessible.

The quantitative standards include:

The Basel framework does not specify the analytical approaches for operational risk, but a bank must show that its adopted measure for operational risk meets a soundness standard such as a one year holding period and a 99.9^{th} percentile confidence interval.

Detailed criteria for calculating operational risk capital include:

1. Any operational risk measurement system must be consistent with the Basel Committee definition of operational risk and the loss event types detailed in Tables 3 through 9.

Table 3: Definition of Event Type Internal Fraud

Event-Type Category (Level 1)	**Definition**
Internal fraud	Losses due to acts of a type intended to defraud, misappropriate property or circumvent regulations, the law or company policy, excluding diversity/ discrimination events, which involves at least one internal party.
Categories (Level 2)	**Activity Examples (Level 3)**
Unauthorized Activity	Transactions not reported (intentional) Transaction type unauthorized (w/monetary loss) Mismarking of position (intentional)
Theft and Fraud	Fraud / credit fraud / worthless deposits Theft / extortion / embezzlement / robbery Misappropriation of assets Malicious destruction of assets Forgery Check kiting Smuggling Account take-over / impersonation / etc. Tax non-compliance / evasion (wilful) Bribes / kickbacks Insider trading (not on firm's account)

Table 4: Definition of Event Type External Fraud

Event-Type Category (Level 1)	**Definition**
External fraud	Losses due to acts of a type intended to defraud, misappropriate property or circumvent the law, by a third party.
Categories (Level 2)	**Activity Examples (Level 3)**
Theft and Fraud	Theft/Robbery Forgery Check kiting
Systems Security	Hacking damage Theft of information (w/monetary loss)

Table 5: Definition of Event Type Employment Practices & Workplace Safety

Event-Type Category (Level 1)	**Definition**
Employment Practices & Workplace Safety	Losses arising from acts inconsistent with employment, health or safety laws or agreements, from payment of personal injury claims, or from diversity / discrimination events.
Categories (Level 2)	**Activity Examples (Level 3)**
Employee Relations	Compensation, benefit, termination issues Organized labor activity
Safe Environment	General liability (slip and fall, etc.) Employee health & safety rules events Workers compensation
Diversity & Discrimination	All discrimination types

Table 6: Definition of Event Type Clients, Products & Business Practices

Event-Type Category (Level 1)	**Definition**
Clients, Products & Business Practices	Losses arising from an unintentional or negligent failure to meet a professional obligation to specific clients (including fiduciary and suitability requirements), or from the nature or design of a product.
Categories (Level 2)	**Activity Examples (Level 3)**
Suitability, Disclosure & Fiduciary	Fiduciary breaches / guideline violations Suitability / disclosure issues (KYC, etc.) Retail customer disclosure violations Breach of privacy Aggressive sales Account churning Misuse of confidential information Lender liability
Improper Business or Market Practices	Antitrust Improper trade / market practices Market manipulation Insider trading (on firm's account) Unlicensed activity Money laundering
Product Flaws	Product defects (unauthorized, etc.) Model errors
Selection, Sponsorship & Exposure	Failure to investigate client per guidelines Exceeding client exposure limits
Advisory Activities	Disputes over performance of advisory activities

Table 7: Definition of Event Type Damage to Physical Assets

Event-Type Category (Level 1)	**Definition**
Damage to Physical Assets	Losses arising from loss or damage to physical assets from natural disaster or other events.
Categories (Level 2)	**Activity Examples (Level 3)**
Disasters and other events	Natural disaster losses Human losses from external sources (terrorism, vandalism)

Table 8: Definition of Event Type Business Disruption and System Failures

Event-Type Category (Level 1)	**Definition**
Business Disruption and System Failures	Losses arising from disruption of business or system failures.
Categories (Level 2)	**Activity Examples (Level 3)**
Systems	Hardware Software Telecommunications Utility outage / disruptions

Table 9: Definition of Event Type Execution Delivery & Process Management

Event-Type Category (Level 1)	**Definition**
Execution, Delivery & Process Management	Losses from failed transaction processing or process management, from relations with trade counterparties and vendors.
Categories (Level 2)	**Activity Examples (Level 3)**
Transaction Capture, Execution & Maintenance	Miscommunication Data entry, maintenance or loading error Missed deadline or responsibility Model / system misoperation / Other task misperformance Accounting error / entity attribution error Delivery failure / Collateral management failure Reference Data Maintenance
Monitoring and Reporting	Failed mandatory reporting obligation Inaccurate external report (loss incurred)
Customer Intake and Documentation	Client permissions / disclaimers missing Legal documents missing / incomplete
Customer / Client Account Management	Unapproved access given to accounts Incorrect client records (loss incurred) Negligent loss or damage of client assets
Trade Counterparties	Non-client counterparty misperformance Misc. non-client counterparty disputes
Vendors & Suppliers	Outsourcing Vendor disputes

2. The bank is required to calculate the capital requirement as the sum of expected and unexpected losses unless it is able to prove that it has measured and accounted for the expected losses exposure.
3. The measurement system adopted by the bank must be able to capture severe tail loss events and risk measures for different operational risk estimates must be added.
4. To meet the soundness standard, any operational risk measurement system must include the use of internal data, relevant external data, scenario analysis, and factors reflecting the business environment and internal control systems. Banks must also have a credible, transparent, well-documented, and verifiable approach for weighting these elements.
    - Internal Data: the development of a reliable operational risk measurement system depends on the tracking of internal loss event data which is needed to match the bank risk estimates to its actual loss experience. A minimum five-year observation period of internal loss data is needed for internally generated operational risk measures. When the bank first moves to the AMA, a three-year historical data window is acceptable. The committee sets standards for a bank's internal loss collection processes.
    - External Data: when a bank is exposed to severe losses, relevant external data must be used with information on the scale of business operations where the event occurred, information on the causes and circumstances of the loss events, or other information that would help in assessing the relevance of the loss event for other banks. The conditions for external data use must be reviewed periodically.
    - Scenario Analysis: is used in combination with external data to assess exposure to high severity events and the potential losses resulting from multiple concurrent operational risk loss events. This approach relies on the knowledge of experienced business managers and risk management experts to derive reasoned assessments of plausible severe losses.
    - Business Environment and Internal Control Factors: must be addressed in order to align capital assessments with risk management objectives and to directly identify improvements or declines in operational risk profiles. This directly reflects the quality of the bank's control and operating environments. The committee sets standards for the use of these factors in a bank's risk measurement framework.

Under the AMA, the recognition of insurance mitigation is limited to 20 % of the total operational risk capital charge calculated. The ability of a bank to utilize such risk mitigation is subject to compliance with criteria set by the committee.

II. Principles for the Sound Management of Operational Risk

In order to improve the effectiveness of operational risk management in the banking industry, the BCBS (2011) set out principles for the sound management of operational risk. These principles reflect the effectiveness of the board of directors and senior management in overseeing its portfolio of products, activities, processes, and systems.

The Committee developed a set of eleven principles distributed as follows:

1. The Fundamental Principles of Operational Risk Management

"Principle 1: the board of directors should take the lead in establishing a strong risk management culture. The board of directors and senior management should establish a corporate culture that is guided by strong risk management and that supports and provides appropriate standards and incentives for professional and responsible behavior. In this regard, it is the responsibility of the board of directors to ensure that a strong operational risk management culture exists throughout the whole organization".

"Principle 2: banks should develop, implement and maintain a Framework that is fully integrated into the bank's overall risk management processes. The Framework for operational risk management chosen by an individual bank will depend on a range of factors, including its nature, size, complexity and risk profile".

2. Governance

   a. The Board of Directors

"Principle 3: the board of directors should establish, approve and periodically review the Framework. The board of directors should oversee senior management to ensure that the policies, processes and systems are implemented effectively at all decision levels".

"Principle 4: the board of directors should approve and review a risk appetite and tolerance statement for operational risk that articulates the nature, types and levels of operational risk that the bank is willing to assume".

b. Senior Management

"Principle 5: senior management should develop for approval by the board of directors a clear, effective and robust governance structure with well defined, transparent and consistent lines of responsibility. Senior management is responsible for consistently implementing and maintaining throughout the organization policies, processes and systems for managing operational risk in all of the bank's material products, activities, processes and systems consistent with the risk appetite and tolerance".

3. Risk Management Environment

a. Identification and Assessment

"Principle 6: senior management should ensure the identification and assessment of the operational risk inherent in all material products, activities, processes and systems to make sure the inherent risks and incentives are well understood".

"Principle 7: senior management should ensure that there is an approval process for all new products, activities, processes and systems that fully assesses operational risk".

b. Monitoring and Reporting

"Principle 8: senior management should implement a process to regularly monitor operational risk profiles and material exposures to losses. Appropriate reporting mechanisms should be in place at the board, senior management, and business line levels that support proactive management of operational risk".

c. Control and Mitigation

"Principle 9: banks should have a strong control environment that utilizes policies, processes and systems; appropriate internal controls; and appropriate risk mitigation and/or transfer strategies".

4. Business Resiliency and Continuity

"Principle 10: banks should have business resiliency and continuity plans in place to ensure an ability to operate on an ongoing basis and limit losses in the event of severe business disruption".

5. Role of Disclosure

"Principle 11: a bank's public disclosures should allow stakeholders to assess its approach to operational risk management".

2.1.3 The Basel III Framework

Basel III is a 2009 international regulatory accord that introduced a set of reforms designed to mitigate risk within the international banking sector. It requires banks to maintain proper leverage ratios and keep certain levels of reserve capital on hand. Basel III framework, with its finalized post-crisis reforms, was a direct response to the economic crisis of 2008. The current date of Basel III implementation is effectively 1st of January 2023.

Under Basel III framework, banks must meet the following capital requirements:

- Common Equity Tier 1 must be at least 4.5% of risk-weighted assets at all times.
- Tier 1 capital must be at least 6.0% of risk-weighted assets at all times.
- Total Capital (Tiers 1 & 2 capital) must be at least 8.0% of risk-weighted assets at all times

Basel III framework delivers the New Standardized Approach for measuring the minimum capital requirements for operational risk. This framework replaces all existing approaches in Basel II framework.

This method is based on the following components:

1. "Business Indicator (BI), which is a financial-statement-based proxy for operational risk".
2. "Business Indicator Component (BIC), which is calculated by multiplying the BI by a set of regulatory determined marginal coefficients (α_i)".
3. "Internal Loss Multiplier (ILM), which is a scaling factor that is based on a bank's average historical losses and the BIC".

<u>Business Indicator</u>

The business indicator consists of

- The interests, leases, and dividend component (ILDC)
- The services component (SC)
- The financial component (FC)

Thus, the BI is calculated as: BI = ILDC + SC + FC

$$\text{BI} = \text{Min}\ [\overline{\text{Abs(Interest Income} - \text{Interest Expense)}};\ 2.25\%\ \overline{\text{Interest Earning Assets}}] + \overline{\text{Dividend Income}}$$

$$\text{SC} = \text{Max}\ [\overline{\text{Other Operating Income}};\ \overline{\text{Other Operating Expense}}] + \text{Max}\ [\overline{\text{Fee Income}};\ \overline{\text{Fee Expense}}]$$

$$\text{FC} = \overline{\text{Abs (Net P \& L Trading Book)}} + \overline{\text{Abs (Net P \& L Banking Book)}}$$

The terms in the formulas above are calculated as the average over three years: t, t – 1, and t – 2, where the absolute value of net items must be calculated first year by year and then averaged over the three years. Tables 10 and 11 summarize the components of the business indicator.

Table 10: Business Indicator Components ILDC and FC

BI Component: Interests, Leases, and Dividend Component		
P&L	Description	Typical sub-items
Interest income	Interest income from all financial assets and other interest income (includes interest income from financial and operating leases and profits from leased assets)	• Interest income from loans and advances, assets available for sale, assets held to maturity, trading assets, financial leases and operational leases • Interest income from hedge accounting derivatives • Other interest income • Profits from leased assets
Interest expenses	Interest expenses from all financial liabilities and other interest expenses (includes interest expense from financial and operating leases, losses, depreciation and impairment of operating leased assets)	• Interest expenses from deposits, debt securities issued, financial leases, and operating leases • Interest expenses from hedge accounting derivatives • Other interest expenses • Losses from leased assets • Depreciation and impairment of operating leased assets
Interest earning assets	Total gross outstanding loans, advances, interest bearing securities (including government bonds), and lease assets measured at the end of each financial year	
Dividend income	Dividend income from investments in stocks and funds not consolidated in the bank's financial statements, including dividend income from non-consolidated subsidiaries, associates and joint ventures.	
BI Component: Financial		
Net profit (loss) on the trading book	• Net profit/loss on trading assets and trading liabilities (derivatives, debt securities, equity securities, loans and advances, short positions, other assets and liabilities) • Net profit/loss from hedge accounting • Net profit/loss from exchange differences	
Net profit (loss) on the banking book	• Net profit/loss on financial assets and liabilities measured at fair value through profit and loss • Realized gains/losses on financial assets and liabilities not measured at fair value through profit and loss (loans and advances, assets available for sale, assets held to maturity, financial liabilities measured at amortized cost) • Net profit/loss from hedge accounting • Net profit/loss from exchange differences	

Table 11: Business Indicator Component SC

BI Component: Services		
P&L	Description	Typical sub-items
Fee & commission income	Income received from providing advice and services. Includes income received by the bank as an outsourcer of financial services.	Fee and commission income from: • Securities (issuance, origination, reception, transmission, execution of orders on behalf of customers) • Clearing and settlement; Asset management; Custody; Fiduciary transactions; Payment services; Structured finance; Servicing of securitizations; Loan commitments and guarantees given; and foreign transactions
Fee & commission expenses	Expenses paid for receiving advice and services. Includes outsourcing fees paid by the bank for the supply of financial services, but not outsourcing fees paid for the supply of non-financial services (e.g. logistical, IT, human resources)	Fee and commission expenses from: • Clearing and settlement; Custody; Servicing of securitizations; Loan commitments and guarantees received; and Foreign transactions
Other operating income	Income from ordinary banking operations not included in other BI items but of similar nature (income from operating leases should be excluded)	• Rental income from investment properties • Gains from non-current assets and disposal groups classified as held for sale not qualifying as discontinued operations (IFRS 5.37)
Other operating expenses	Expenses and losses from ordinary banking operations not included in other BI items but of similar nature and from operational loss events (expenses from operating leases should be excluded)	• Losses from non-current assets and disposal groups classified as held for sale not qualifying as discontinued operations (IFRS 5.37) • Losses incurred as a consequence of operational loss events (eg fines, penalties, settlements, replacement cost of damaged assets), which have not been provisioned/reserved for in previous years • Expenses related to establishing provisions/reserves for operational loss events

Business Indicator Component

The BIC is the product of the BI and the marginal coefficients. Table 12 shows these parameters for each of the three buckets. The marginal coefficients increase with the size of the BI.

Table 12: BI Range and Coefficients

Bucket	BI Range (€ bn)	BI Marginal Coefficients (α_i)
1	≤ 1	12 %
2	$1 < BI \leq 30$	15 %
3	> 30	18 %

Internal Loss Multiplier

$$ILM = \ln\left(\exp(1) - 1 + (LC/BIC)^{0.8}\right)$$

The operational risk capital calculation is affected by the internal loss multiplier which represents the bank's internal operational loss experience. The Loss Component (LC) is equal to 15 times the average annual operational risk losses incurred over the previous 10 years.

If the LC and BIC are equal, then the ILM is equal to 1. While, if the LC is greater than the BIC then the ILM is greater than 1 which is the case of a bank with losses that are high relative to its BIC. Thus, it is required to hold higher capital due to the incorporation of internal losses into the calculation methodology.

On the contrary, if the LC is smaller than the BIC, then the ILM is less than one which is the case of a bank with losses that are low relative to its BIC. Thus, it is required to hold lower capital due to the incorporation of internal losses into the calculation methodology.

The calculation of average losses in the Loss Component must be based on 10 years of high-quality annual loss data, the collection of which is subject to qualitative requirements. However, banks that do not have 10 years of high-quality loss data may use a minimum of five years of data to calculate the Loss Component (this exception does not apply to banks which are currently using the AMA).

Minimum Operational Risk Capital

Under the Standardized Approach, the operational risk capital is the product of the BIC and ILM.

$$ORC = BIC \times ILM$$

For banks in bucket 1, the internal loss data does not affect the capital determination since the ILM is equal to 1 and thus ORC = BIC = 12 % × BI.

General Criteria for the Loss Component

- Banks must have documented procedures and processes for the identification, collection, and treatment of internal loss data which must be linked to current business activities, technological processes and risk management procedures.
- Banks' internal loss data must be all-inclusive and must capture all material activities and exposures from all applicable subsystems and locations. A €20,000 minimum threshold for including a loss event in the data collection and calculation of average annual losses is set.
- Banks must collect the following information regarding operational risk events: date of occurrence, date of discovery, and the date (or dates) when a loss event results in a loss, reserve, or provision against a loss being recognized in the bank's profit and loss (P&L) known as date of accounting. Banks must also collect information on recoveries of gross loss amounts and the drivers or causes of the loss event.
- Operational loss events that relate to credit risk, but are not accounted for in credit risk RWAs should be included in the loss data set.
- Operational risk losses related to market risk are treated as operational risk for the purposes of calculating minimum regulatory capital under this framework and will therefore be subject to the Standardized Approach for operational risk.

2.2 Research in Operational Risk

Embrechts et al., (2003) examine the VaR approach for calculating capital requirements for operational risk under Basel II. The capital charge is the summation of the VaR at confidence level α of each of the different business lines required by the BCBS. VaR can be defined as "a statistical estimation of a portfolio loss with the property that, with a given (small) probability, we stand to incur that loss or more over a given (typically short) holding period". It is important to note that VaR techniques become delicate with confidence levels of 99.9% and beyond (as in operational risk) since there is barely repetitive data to predict the losses of such magnitude. VaR estimates can be derived by combining historical simulation and Extreme Value Theory techniques. Since operational risk measurement involves extreme loss events, the authors' emphasis was directed to the assumptions underlying EVT. The authors use findings of a simulation study which compares the estimated quantiles with the corresponding theoretical ones for known distributions for which (high) quantiles can be calculated explicitly with datasets of {25, 50, 100, 200} exceedances. Three types of loss distributions are applied which are medium-tailed, heavy-tailed with infinite moments of order greater than or equal to two, and heavy-tailed with infinite moments of order greater than or equal to one. Thus, the Lognormal Distribution and Pareto Distribution with $\theta = 2$ and $\theta = 1$ are adopted respectively. The results show that larger sample sizes are needed with the heavier tails to obtain the desired accuracy. For a total number of 287 losses over a one year period, if the loss data are of the first type, i.e. LD, then estimating the VaR at the 99.9% level with 287 observed data points could be justified. While, if the loss data are of the second class type, i.e. PRD $\theta = 2$, then the VaR at 99% confidence level can be sufficiently estimated with the targeted accuracy. However, for a 99.9% confidence level, 287 observations are not enough and 670 loss values or more are needed. Finally, considering that the loss data are of the third class type, i.e. PRD $\theta = 1$, 287 observations, assumed to be iid and repetitive, are even not enough to estimate the VaR at the 99% confidence level. The estimation of high quantiles is an inherently difficult problem. The authors highlight that the best way to gain control over operational risk is to increase the quality of control over the possible sources of huge operational losses because the latter cannot be considered as simple accidents. Thus, the authors conclude that Pillar 2 and in part Pillar 3 of Basel II are extremely important, and that Pillar 1 should not be overemphasized.

Couto and Bulhoes, (2008) apply the BIA, the SA, and the ASA to quantify operational risk capital charge for seven eminent financial institutions in Portugal. Gross income values were extracted from the annual bank reports for 2002 and until 2006. The authors note that the AMA of Basel II was not applied due to the impossibility of retraction of internal data on operational losses. For the ASA, loans and advances for retail and commercial banking were also extracted. Results show that the SA generates a decrease in the capital required in all seven institutions as compared to the BIA. Thus, the progression of a bank from BIA to SA under Basel II is recommended. Whereas for ASA, the results were variable, some institutions witnessed a decrease in the capital requirement while others showed an increase. The progression directly from BIA to ASA generates significant reductions in capital charge for most of the institutions. Consist with the literature, the authors highlight the benefits of progressing from BIA to SA, yet reservations are noted regarding the factors assigned by Basel II for each business line. In addition, the authors note that the transition directly from BIA to ASA is advantageous in most situations. Finally, the authors emphasize the obstacles involved in analyzing operational risk including the correct quantification of losses, the inclusion of all operational risk situations, and the relevance of recorded risks over time.

Moosa, (2008) criticizes the AMA proposed by the BCBS and considers it as problematic. There is no actual agreement of what constitutes this approach. The AMA requires banks to build their own internal models to calculate the capital charge for operational risk, and the motivation provided by the Committee for banks to adopt the AMA is that it yields lower capital charges as compared to the BIA or the SA. Moosa discourages banks to adopt the AMA since he considers that there is no understandable reason it would produce lower capital charges. The author addresses his critique of the AMA first by examining what constitutes this approach. There is no clear list of techniques which fall under the AMA. The BCBS proposes the Loss Distribution Approach, the Internal Measurement Approach, the Scorecard Approach, and the Scenario Based Approaches. However, authors argue that the AMA is not restricted to the mentioned techniques, as it could be based on any technique which leads to the precise measurement of operational risk exposure. The AMA provides flexibility for banks to develop their own internal models, but this leaves the approach ambiguous. Moosa also tackles the issue of whether the proposed techniques under the AMA are treated as independent and considered alternatives or they must be used jointly and are considered as complementary. The Committee, in 2003, considers that the SCA and the SBA can be used to supplement the data used by the LDA or that they can be used separately. However, in 2006 the Committee considers that all techniques must be used jointly and that the bank's approach for weighting the four fundamental elements (internal data, external data, scenario analysis, and business environment and internal control factors) should be internally consistent. Furthermore, Moosa considers that the AMA is not viable in terms of costs and benefits and that it has brought many complaints from banks because it is too expensive and complex.

Esterhuysen et al., (2008) calculate regulatory capital under the AMA based on a VaR model using actual operational loss data from a retail bank in South Africa. The LDA was used to calculate the VaR. The data used, which is regarded as internal loss data, includes the operational losses and the gross income for the last three years. The SA was also adopted in calculating the capital charge. Since the data used is that of a retail bank, thus only one business line (retail banking) is applied. Accordingly, the regulatory capital using the SA is calculated as 50,891,219 R. Under the AMA, the Poisson distribution is used to model the frequency, while the Exponential Distribution is used to model the severity. The aggregate loss distribution is determined through a Monte Carlo simulation. Finally, the percentiles are calculated in order to determine the VaR. For a 99.9% confidence level, as per Basel II requirements, the capital charge is calculated as 13,384,748 R. The authors note the big reduction in operational risk capital requirement when shifting from the SA to the AMA. They relate this to the fact that the SA is based on gross income which is a poor indication of the real operational risks occurring in a bank. The capital value under the AMA will be a truer reflection of the operational risk a bank faces.

Gregoriou, (2009) classifies internal models developed by banks under the AMA to two categories, the LDA and the SBA. The LDA is a parametric method which is based on previous observed internal loss data possibly supplemented with external data. The LDA involves an estimation of the frequency distribution for the occurrence of operational losses in addition to a severity distribution for the economic impact of the individual losses. The two distributions are combined using n-convolution of the severity distribution with itself, and n is a random variable that follows the frequency distribution. The main obstacle for developing LDA models is the availability of loss data points, particularly large ones. Financial institutions are somewhat exempt of this issue since its activities include a lot of interaction with other parties and thus operational errors are usually identified. In many cases, such as new business environments, small banks, and specific operational risk classes, employing a comprehensive LDA model would not be feasible due to the difficulty in obtaining enough relevant data points. In such cases SBA under the AMA are adopted, which are based on experts' opinions. Similar to the LDA, the SBA combines frequency and severity to quantify the aggregate loss distribution however the difference is that the estimation is based on experts' about different scenarios. These scenarios are determined first by describing possible adverse events realizing operational risks, this is done by banks. Then, experts provide their opinions on the frequency and severity of the events described. Thus, the chief obstacle is the reliability of the experts' estimation. It is essential that the SBA yield statistically compatible results with the LDA. The most suitable technique to combine LDA and SBA under AMA is the Bayesian inference in which the experts set parameters of the loss distribution.

Angela et al., (2009) apply the LDA to quantify operational risk in a financial institution (bank or insurance company) using empirical observations from OpData dataset of OpVantage (division of Fitch Risk Management). The Expectation and Maximization algorithm is implemented on the dataset to overcome the fact that only operational risk losses exceeding \$1 million are recorded. The database includes loss events from worldwide firms corresponding to the financial services sector. This analysis considers the period from 1994 till 2006 with a total of 1,025 records. Losses are categorized into seven event types as per the Basel II Accord. However, only five event types are considered due to lack of data in business disruption and system failures, and damage to physical assets event types. The PD is used to estimate the frequency of loss, while for severity the LD is used in the left tail and center, whereas the Generalized Pareto distribution is used for the right tail. For each event type, the Monte Carlo simulation is used to estimate the aggregate loss distribution while Value at Risk and Expected Shortfall are used to quantify operational risk capital charge. Capital charges are calculated for each event type at confidence levels 95%, 99%, and 99.9%. The total operational risk capital is quantified in three cases, where in the first case perfect dependence is assumed while in the second, independence between two event types is assumed. In the third case dependence structure between event types is modeled using a t-Student copula. The results, expressed as a percentage of total assets, show that the application of the t-Student copula produces capital charges as expected. That is, the values fall between the minimum (assumption of independence) and the maximum (assumption of perfect dependence). Thus, the authors note the importance of considering the dependence between event types. Furthermore, since the results were presented as a percentage of the total assets at different confidence levels, this allows financial institutions to use these results to measure their own operational risk capital charges each based on its total assets. Finally, the authors highlight the importance of the quality of operational loss databases and its impact on the final results. The authors recommend that a financial institution should not rely on its internal data only, yet it must integrate with good quality databases.

Momen et al., (2012) model the operational risk of an Iranian private bank based on the LDA. The authors highlight that Iranian banks suffer from loss data unavailability, unreported large losses, and lack of attention to dependence structure of operational risks. Starting with frequency distribution, the Negative Binomial Distribution was selected for all risk categories based on the results of three statistic hypothesis tests. While regarding severity distribution, discrimination between ordinary (high frequency and low impact) and large (low frequency and high impact) losses has been considered. Furthermore, SCA was implemented to model the tail of the severity distribution since large disastrous losses such as bankruptcy have not been reported in Iranian banks. In this approach, bank experts provide scenarios about frequency and severity of large losses in 1 year. As per the Basel II Accord, a combination of the eight business lines and the seven event types was adopted thus generating a 56-cell matrix. However, due to the scarcity of data, only four cells were used for modeling in this study covering two business lines which are Retail and Commercial Banking with two event types which are Business Disruption and System Failures and Execution, Delivery, and Process Management. Monte Carlo simulation was used to estimate the aggregate operational losses for each cell. Finally, dependence among aggregate losses is modeled by a multidimensional t-copula. The Basel Committee has assumed a perfect positive dependence between risks since it mentions calculating the total capital charge of the bank by simple summation of the capital charges of all 56 risk categories. However, banks are interested in considering the dependence structure to yield a lower capital charge. Techniques used in modeling dependence between variables include correlation, yet studies show the superiority of copula over correlation for modeling dependence due to higher flexibility of the copula and its ability in modeling dependence between extreme events which are the central concern in operational risk modeling. The results modeled with t copula and a 99.9% confidence level generated a Capital at Risk of Karafarin Bank almost equal to 7.4×10^{11} IRR. As compared with other approaches under Basel II, such as the BIA or the SA, the model requires much less capital which was consistent with the bank's management presumptions. Thus, using the model allows banks to use the extra unallocated capital for creating further income within a controlled level of operational risk. The authors recommend considering other multivariate copulas and modeling operational risk with other methods under the AMA such as Bayesian approaches, neural networks, or Fuzzy modeling.

Pietro et al., (2012) apply the double transformation kernel estimation to approximate the severity distribution of operational risk. Noting the difficulties in estimating the severity distribution and its significance on the capital requirement, the non-parametric kernel estimation is proposed as an alternative to parametric estimations. The analysis is based on double transformation to improve the severity estimation, and the operational losses data sample of a medium-sized Spanish Savings bank is used. Double transformation kernel estimation involves carrying out an initial processing of the data by using the distribution function of the generalized Champernowne with three parameters. Thus, the transformed variable has a distribution that is positioned around a Uniform distribution (0, 1). Successively, the data is transformed again by using the inverse of the Beta function (3, 3). Finally, the result of this double transformation will have a distribution close to that of Beta (3, 3). The data used is internal loss data in retail banking ranging from 2004 to 2006. Parametric approaches in estimating the severity distribution are applied for comparison purposes, namely Weibull Distribution, LD, and GPD. The Cramer-Von Mises (CVM) and the Kolmogorov-Smirnov (KS) goodness-of-fit tests are implemented, the results of which show that the distribution of the double transformation kernel estimation provides the best fit to the data. Next, the severity distributions obtained are aggregated with a PD to determine the operational VaR. Results show that VaR approximated using parametric approaches tends to be overestimated or underestimated, while the double transformation kernel estimation presents the more realistic estimation compared with the empirical estimation. Thus, the authors highlight the use of the double transformation kernel estimation as a substitute for the estimation of operational risk loss severity distribution as it comprises all tail behavior and includes data of high-frequency small losses which form the body of the distribution.

Lin et al., (2013) apply the BIA, the SA, and the AMA to quantify the operational risk capital requirement of a bank in Taiwan with all of its branches and headquarters. Loss data of the bank used in this study ranges from January 2004 to December 2006 with a total of 320 records and total loss amount of $16.9 million. Since the Basel Committee requires data over 3 years for banks using AMA for the first time, EVT is used on the operational loss data of one case bank to examine the operational VaR. Results show that the capital charge calculated by the SA is $134.01million which is less than that of the BIA, $142.1 million. Under the AMA, EVT is used to estimate the VaR of the operational loss using the Peak over Threshold method. The GPD is used in which any value exceeding the threshold meets the distribution. The threshold is determined from the loss curve as $64,516.13. The Maximum Likelihood Method and the Least Square Method are used to estimate the parameters of the GPD, the LD, and the PD. The parameters are imported into R platform and 100,000-time stimulations are conducted, and the loss exceeding probability (EP) at a 99.6% confidence level is estimated. Different results are obtained for the estimation models used, however the highest capital charge is determined as $16.7 million using the PD at 99.9% confidence level which is drastically less than that of the BIA and SA. Thus, the implementation of the EVT to estimate the operational VaR and consequently to calculate the operational risk capital charge for banks will decrease the regulatory capital required for banks to hold in large amount. The authors note that within the banking industry, the BIA and SA are generally employed, and that there has been no researcher adopting the AMA. However, the adoption of the GPD or PD model of the EVT under the AMA results in a radically less amount in operational risk capital charge.

Corrigan et al., (2013) investigate the current state of operational risk assessment frameworks and their development to meet the emerging needs of the future. The chief research findings include firstly that operational risk is a material risk and it is one of the main reasons for organizational failure. It was with Basel II, in 2004, that banks started quantifying their operational risk and insurance companies began to follow in the same path. The value of an effective operational risk management framework is increasing in the development of resilient organizations. Secondly, it is important to note the nature of operational risks which are highly skewed. Even though the number of operational risk loss events are characterized with low severity, however total operational losses are dominated by high-severity and low-frequency events. Thus, any operational risk framework must accommodate for the aforementioned loss event types. Thirdly, the authors note advantages of scenario and statistical modelling approaches such as the LDA over the basic indicators and standard formulas which are blunt tools. However, improvement is required in these approaches to respond to changing operational risk levels during financial crises. Fourthly, the leading models in operational risk are structural or casual-based models. For instance, the Bayesian network is able to flow information in both directions through the Bayesian inference and thus allows for the vigorous determination of operational risk limits which would be consistent with operational risk levels. In addition the authors address the use of phylogenetic techniques in assessing operational risk events. Phylogenetic techniques involve the objective assessment of the relationships of the characteristics of operational risk and facilitate a structured way for including new and emerging risks. Finally, the authors discuss the divergence of regulatory frameworks in assessing operational risk and quantifying the corresponding capital charge and highlight the value of establishing a Loss Data Collections process.

Shevchenko and Peters, (2013) review proposed methods for combining different data sources as a regulatory constraint of the LDA under the AMA of Basel II. The LDA requires banks to quantify distributions for frequency and severity of operational risk losses for each cell of the 56-cell matrix combining the eight business lines and seven event types highlighted in Basel II Accord. One of the criteria which banks using AMA must satisfy includes the use of internal data, relevant external data, scenario analysis through expert opinions, and factors reflecting the business environment and internal control systems. Yet, the main obstacle remains the lack of quality data, since in the past banks did not collect operational risk loss data. Furthermore, the challenge is in combining the different data sources which is addressed in this work. Starting with internal data, it must be collected over a period of five years, but a three year period is accepted for banks starting the AMA. The data must be mapped over all the cells of the 56-cell matrix. Also, a reporting threshold of the order of 10,000 Euros can be adopted. For internal data, requirements include information on gross loss amounts and any recoveries, date of the event, and information regarding the drivers of the loss event. In regards to external data, it must include actual loss amounts, information on the scale of business operations where the event occurred, and information on the causes and circumstances of the loss events. External databases provide industry data, however the data suffer from a survival bias since the data of collapsed companies are not included. Thirdly, for scenario analysis, Basel II requires banks to use scenario analysis in combination with internal and external data to assess exposure to high severity events. Scenario analysis is very subjective since it involves opinions of experienced business managers and risk management experts in the identification and analysis of risks. Most commonly, opinions are collected on distribution parameters, the number of losses with the amount to be within some ranges, the frequency of the losses and quantiles of the severity, and how often the loss exceeding some level may occur. However, expert elicitation remains one of the main challenges in operational risk due to the various backgrounds of managers which might create some misunderstandings. Additionally, another problem is the fact that scenario analysis is done at the loss process level while external data is collected for the risk cells of the matrix, this creates problems in data combining. Furthermore, since Basel II requires that the risk measure used for capital charge should correspond to the 99.9% confidence level for a one-year holding period, data sufficiency is another aspect which must be considered. The amounts of data on annual losses needed to meet the aforementioned requirement are not available, even from the largest databases. As a result, parametric models must be adopted. Finally, factors reflecting the business environment and internal control systems must be incorporated as well. Thus, the major obstacle lies in combining data from the different sources discussed above for model estimation. The proposed methods include Ad-hoc combining, parametric and nonparametric Bayesian methods, and general non-probabilistic methods such as Dempster-Shafer theory. Ad-hoc combining is primarily used to combine internal data, external data and expert opinions. The minimum variance estimator theorem is applied. Assuming that the estimators are unbiased is considered reasonable when combining estimators from different experts, or from expert and internal data. However, it is doubtful when applied to combine estimators from the external and internal data. Regarding Bayesian methods, the Bayesian inference method can be used to combine different data sources in a consistent statistical framework. Moreover, nonparametric Bayesian approaches could be adopted such as the Dirichlet process which represents a probability distribution of the probability distributions. Lastly, the Dempster-Shafer theory is proposed as a method for combining data from different sources particularly in situations where there is little information on which to evaluate a probability or when information is ambiguous or conflicting. Researchers highlight that the use of p-boxes and Dempster-Shafer structures in risk analyses offers many significant advantages over a traditional probabilistic approach. However, it might be problematic to justify application of Dempster's rule to combine statistical bounds for empirical data distribution with exact bounds for expert opinions. The authors concludes that Bayesian framework is suitable to combine different data sources such as internal data, external data and expert opinions, and to account for the relevant uncertainties.

Lu et al., (2013) implement the Semi-linear credibility theory in operational risk measurement using Chinese commercial banks' data from 1990 to 2011. Credibility theory is widely employed in non-life insurance. It uses the previous loss information of an insured to estimate the cost of providing references for future pricing. Credibility theory was proposed by Buhlmann and it could be adopted by banks to calculate operational risk capital requirements. Semi-linear credibility theory is needed to transform large claims before integrating the original and transformed data in estimating the credibility premiums of the next period. That is, because when large claims occur, the credibility weights are rather small for small and medium insurance contracts thus credibility premiums will vary which is inconsistent. As a first step, a threshold must be selected as a truncation point since the operational loss data is characterized by a fat tail. The Hill plot, mean excess plot, and kurtosis methods are used to determine the threshold value. Moreover, two threshold values are selected empirically to improve accuracy. The Chi-square test of goodness of fit is implemented to select the best threshold value. The MLE is used to generate the parameters of the POT model. An evaluation of the six groups of banks demonstrates that the estimations differ for the tail of the banks' internal data. Results indicate that the operational risk capital reserve must be greater than the mean of the one-year operational risk loss for the following year. Thus, the Semi-linear credibility model is applied to integrate the internal and external loss data of the six groups of banks to generate the mean operational risk loss of the six groups of banks separately. Subsequently, the PD is used for frequency distribution and VaR values are generated. Operational risk capital charge is also calculated using different approaches such as the LDA under the AMA, and the BIA. Results show that the Semi-linear credibility model will save approximately Yn 245.5 billion in capital for Chinese commercial banks every year when compared to the AMA, and it is better than the BIA since it will also save between Yn 2.6 billion to Yn 59.2 billion in capital. The authors highlight that the Semi-linear credibility model increases the profitability and enhances the competitiveness of Chinese commercial banks.

Rahman et al., (2014) present the Bayesian inference method for the modeling and estimation of operational risk in a bank. The dataset used belongs to the ET4, clients, products, and business practices where the simulated bank data has 152 internal and external loss datasets. The loss data is categorized over 16 risk cells which are then related to the eight business lines detailed by the Basel Committee. Regarding the estimation of frequency and severity prior, MLE procedure is used to estimate the prior gamma distribution parameters α and β, and the prior normal distribution parameters μ and σ. The prior distribution parameters estimated are then used for the estimation of frequency and severity posterior distributions for each cell. The Poisson frequency parameter λ and log-normal severity parameter μ uncertainties are taken into consideration. The entire ET4 loss data is fitted into a LD to determine a point estimation of the standard deviation which is used in determining the posterior distributions of the parameters in addition to the frequency and severity distributions. Monte Carlo simulation is conducted with a K of 1,000,000 for each of the 16 cells. Regarding the high severity losses, they are broken into six buckets. Only buckets 4 through 6 are considered since they cover the higher loss levels which are \$1M – \$5M in bucket 4, \$5M – \$25M in bucket 5, and losses higher than \$25M are in bucket 6. Buckets 1 through 3 are ignored since those losses are already included by the internal and external data. The bucket frequencies of high severity losses are defined by bank experts each year. The same Monte Carlo simulation is also conducted for this structured scenario. The loss data, internal and structured scenario, are merged and added to get the aggregated total which is then sorted by the level of severity. The capital charge is calculated for a confidence level of 99.9%, 99.93%, and 99.97%. For comparison, the capital charge values are then determined using frequentist methods which do not include the effect of parameter uncertainty while estimating the capital requirement. The results from both approaches, Bayesian inference and frequentist methods, are rather similar with the frequentist approach estimating slightly higher capital charge values. The authors highlight that the main advantage of Bayesian inference is that it takes into consideration the parameter uncertainties.

Zhu et al., (2014) address the severity distribution in the LDA by proposing a nonparametric approach based on Cornish-Fisher expansion for operational risk modeling. Since the choice of severity distribution has a distinct impact on the capital requirement, as compared to the frequency distribution, an inadequate choice would lead to overestimated or underestimated capital charges. In most cases, researchers use parametric approaches to estimate the severity distribution such as the LD, WD, and ED which fit high-frequency and low-severity loss events. However, the tail of the severity distribution, which includes the low-frequency and high-severity losses, cannot be fitted with such distributions. Thus, tail distributions such as the g-h distribution or the α-stable distribution are employed. Furthermore, the EVT is also employed, however incorrect threshold selections for the start of the tail lead to compromised results. Thus, under the LDA, different severity distribution assumptions results in noteworthy differences in the capital charge calculated. The commonly used distributions mentioned above either cannot fit both the body and the tail or cannot determine an objective threshold value. Nonparametric approaches on the contrary, do not assume any particular severity distribution. Accordingly, the Cornish-Fisher expansion is adopted to estimate severities under the LDA framework. Cornish-Fisher expansion is a mathematical expansion which is used to approximate the quantiles of a random variable based on its first few cumulants or moments, it does not require a predetermined distribution to fit the loss severity of operational risk. The standard normal distribution along with the Cornish-Fisher expansion is used to produce samples of severity which are used in a Monte Carlo simulation to sketch the annual operational risk loss distribution. Figure 1 summarizes the procedure of the proposed framework. VaR is used to measure the magnitude of operational risk.

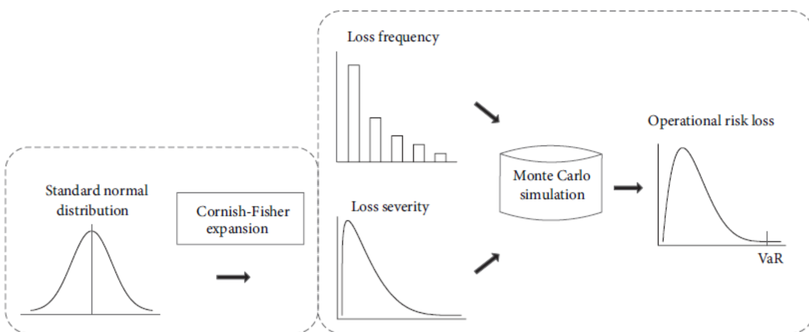


Figure 1: Summary of the LDA Procedure

The data used in this research is from an operational risk database of Chinese banking spanning from 1994 to 2012 with a total of 2132 collections. Each record is manually searched, labelled, and sorted out from public resources, such as the newspapers, the internet, and court documents. The end time and the loss amount are exacted from the database. Regarding frequency distribution, the NBD is selected based on the results of the KS test. After completing the procedure of the framework described above, results show that the VaR at 99.9% ranges from 67 to 13290 billion CNY. It is important to note that the order of Cornish-Fisher expansion radically affects the magnitude of VaR. where the larger the order of Cornish-Fisher expansion, the more accurate the results are. When higher order moments, such as fourth and fifth moments are added in the expansion, VaR converges to about 82 billion CNY. The authors highlight that the lognormal distribution only uses information of the first and second moments, while the suggested approach is able to include the information of high order moments. Thus, the authors conclude that the LDA with the Cornish-Fisher expansion for severity distribution is capable of yielding a more accurate operational risk capital charge.

Vukovic, (2015) demonstrates an operational risk model by implementing the GPD and the Monte Carlo simulation. A random number generator is used, due to the lack of sufficient data, to simulate the events such that extreme events do not occur frequently. The PD is selected to model the frequency, while for severity, the body and the tail are modeled separately. Based on EVT, a threshold μ is determined to separate the body from the tail. Thus, the LD is used to model the severity of the body, while the GPD is used to model that of the tail. Furthermore, the convolution method is used, through Monte Carlo simulation, with the frequency and severity distributions to obtain the annual loss distribution. A high threshold value μ is set where the parameters are estimated on excess over threshold using probability weighted moments (mean excess function). Consequently, the Q-Q plots demonstrate that the tail of the severity distribution is characterized by a GPD. A Student t-copula is used for the overall loss distribution since it is the combination of Lognormal and GPD. R software is used to conduct the convolution method and to generate the overall loss distribution with VaR results. The author highlights that the presented work represents a review of the state of the art techniques and serves as a model for quantifying operational risk in insurance and financial institutions.

Han et al., (2015) apply the POT model to quantify operational risk in Chinese commercial banks using operational loss data from 1995 till 2012. The POT requires the selection of a threshold, which is a critical step, as it impacts the scale and shape parameters of the GPD. An optimal threshold value must be selected by balancing the relationship between bias and variance. The Anderson-Darling statistic is used for threshold testing. The data sample used is composed of 533 operational loss events from Chinese commercial banks during the period of 1995 till 2012. The data shows that internal fraud is the main source of operational risk. Four threshold values are selected using the mean excess plot, and the MLE is used to estimate the corresponding parameters of the GPD. Based on the results of the Anderson-Darling statistic, a threshold value of 75,000 is selected. VaR and Expected Shortfall are calculated using this threshold at confidence levels of 95%, 99%, 99.5% and 99.9%. Consistent with Basel II, a confidence level of 99.9% is selected and the operational risk capital requirement is 13,922 million RMB. The authors conclude that commercial banks need to strengthen their internal control to decrease internal fraud. The authors also highlight the importance of implementing computer technology to discover illegal operations and thus decrease corresponding bank losses in real time.

Bajaj, (2016) estimates the operational risk capital charges (ORCC) of public and private sector banks in India using the BIA and the SA from 2003 till 2013. For the BIA, data of the total period of 10 years is used, while for TSA since 2006 was the first year on segment reporting as per the Reserve Bank of India (RBI) thus data of the total period of 7 years only is used. Collection of data was conducted through the RBI website, Indian Banks Association's Performance Highlights, and annual reports of the banks included in this work. The period from 2003 till 2013 is divided into eight phases under BIA and five phases under SA where each phase covers 3-year positive average gross income. Furthermore, a bank's Capital Adequacy Ratio (CAR) and Tier 1 capital ratio are examined to assess the impact of operational risk capital charges. Results show how ORCC differs based on a bank's size regardless of the risk management system adopted. A comparison between private and public section banks shows that a considerable proportion of the ORCC is held by public sector banks. It is important to note that capital charges are higher under BIA, as compared to SA. Also, variations in business size and bank income levels influence the GI levels and consequently the ORCCs. In regards to CAR and Tier 1 capital, the analysis shows a decline due to the incorporation of additional ORCCs across Indian public and private sectors banks over the years. The author notes that banks which want to use SA must have a well well-defined business line mapping framework to avoid the burden of high beta factors.

Wang et al., (2017) present the LDA with Piecewise-defined Frequency Dependence (LDA-PFD) to calculate the operational risk capital of the entire Chinese Banking system based on the Chinese Operational Loss Database (COLD). Under Basel II and within the AMA, the LDA is the most common. However, the LDA holds many setbacks mainly in the severity distribution component of operational risk as compared to the frequency distribution component which does not have as much of an impact on capital charge. Operational risk loss distribution has a fat-tail characteristic since the loss is minor in most cases yet major in some extreme cases. Thus, loss events due to operational risk are divided to two categories which are the high-frequency low-severity and low-frequency high-severity loss events. Furthermore, as the Basel Committee divides banks' business activities into 8 business lines, dependencies between these lines must be studied which are the frequency dependence, severity dependence, and loss distribution dependence. The dataset used consists of 2132 records ranging from 1994 to 2012 collected from publicly available information sources. The occurrence of operational risk events mostly appears in three business lines which are Retail Banking (1145 times), Commercial Banking (619 times), and Payment and Settlements (263 times). Other business lines are discarded since loss events in them does not exceed 30 times. From the data, the end time, loss amount in CNY, and business line type are used for the three business lines. The aforementioned data is first used to determine a proper threshold which is 200 million CNY based on the mean excess plot. A threshold is selected to divide the data since the frequency and severity distributions of body and tail of operational risk loss data are different. Results also show that the NBD is better for high-frequency low-severity data while the PD is better for low-frequency high-severity data. Moreover, the Skewed Generalized Error Distribution can fit the high-frequency low-severity data well while the Logarithmic Normal distribution fits the low-frequency high-severity data better. Regarding the copula function, results show that the t copula fits the data better for body and tail as compared to the Gaussian copula. Finally, the Monte Carlo simulation is used to calculate the operational risk capital charge for the Chinese banking system at different confidence levels. Results show that the capital charge is 76 billion CNY at 99.9%. The capital charge is also calculated at 99.9 % using the BIA, LDA, loss distribution approach simply considering frequency dependence of the entire data (LDA-FD), and loss distribution approach based on piecewise-defined distribution but not considering dependence (LDA-PD). Results show that the LDA-PFD approach adopted yields the lowest capital requirements for operational risk which is favored by banks. The authors conclude that the LDA-PFD can obtain more accurate results as compared to the LDA, LDA-FD, and LDA-PD since it makes the distribution fitting more accurate and it can capture a more precise correlation structure.

Tharwat et al., (2018) propose an improvement to the LDA using Fuzzy numbers. The authors note that the LDA is the most common under the AMA, yet they highlight that it suffers from numerous setbacks such as problems related to combining data from different sources, uncertainties in the estimated capital due to subjectivity in the parametric model assumption, and its dependence on past conditions which are assumed to extend to the future. Among other reasons, the authors were motivated to improve the LDA through Fuzzy numbers. The proposed improvement is titled Triangular Fuzzy Number-Loss Distribution Approach (TFN-LDA). In an attempt to avoid using specific distribution or assumptions during estimation of the expected operational loss, operational risk frequency and severity can be expressed as triangular fuzzy numbers. Triangular fuzzy number was used such that the triplet [a, b, c] can be viewed as the [minimum, average, maximum] expected loss. For both frequency and severity, the lower (optimistic), average (realistic), and upper (pessimistic) bounds are determined. Consequently, m sequence triangular fuzzy numbers are generated for frequency and severity, each in its respective interval. Extreme events were expressed by allowing frequency and severity to exceed the upper bounds of the interval. Finally, the risk measure VaR was used to estimate the required operational loss capital for the minimum, average, maximum, and extreme cases. Data used is the operational loss data obtained from a South African retail bank recorded in 2003, 2004, and 2005. The results show that the capital requirement falls in the interval [10,397,622 rand and 19,789,549 rand]. The operational risk capital requirement is also calculated for the same dataset in previous work using the BIA, and the LDA under the AMA. Under BIA, the minimum regulatory capital required to cover the operational losses is calculated as 50,891,219 rand for one year, while under the LDA it is calculated as 13,384,748 rand for one year. The PD is used for frequency, while the ED is used for severity under the LDA. A comparison of the results shows that the BIA leads to a much higher capital charge as compared to the LDA and TFN-LDA which are more consistent. The authors highlight the benefits of the TFN-LDA which include more accuracy in estimating the capital charge, less subjectivity in the parametric model assumption, and the ability to study the impact of extreme events easily. Also, the TFN-LDA provides more flexibility to choose the required capital from a specific interval which reflects the occurrence of extreme events.

Wei et al., (2018) provide a comprehensive overview of the worldwide operational Loss Data Collection Exercises (LDCEs) of the four data elements of the AMA which are internal loss, external loss, scenario analysis, and business environment and internal control factors (BEICFs). The relationship between the aforementioned data elements can be seen in Figure 2.

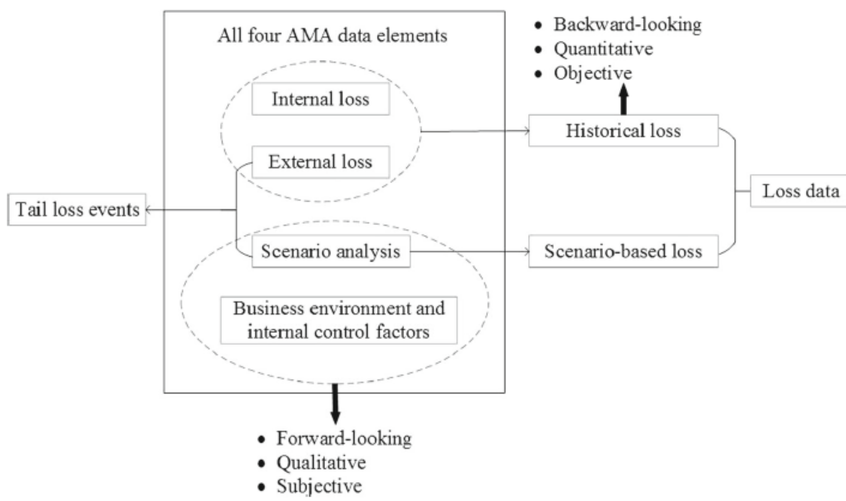


Figure 2: Relationship between Data Elements (Wei et al., 2018)

The authors reviewed a total of 267 papers related to operational risks in banks from 2002 until March 2017, along with 34 relevant articles, and surveyed a large amount of other information. Consequently, the authors classify various sources of operational risk into five types which are individual banks, regulatory authorities, consortia of financial institutions, commercial vendors and, researchers. These sources and the data type from them are summarized in Figure 3.

Data types	Sources				
	Individual banks	Regulatory authorities	Consortia	Commercial vendors	Researchers
Internal loss	√				
External loss					
Pooled industry data		√	√		
Public data			√	√	√
Scenario analysis	√	√	√		√
BEICFs	√	√			

Figure 3: Data Types and Sources (Wei et al., 2018)

Internal loss represents banks' activities, technological processes, and risk management processes which are the most relevant information required for operational risk measurement. It is with Basel II, that many banks started dedicating resources for the collection of loss data and the construction of internal databases. Since the Basel II framework permits banks to set reporting thresholds to collect losses exceeding the threshold amount, this has led to a (left) truncation biased internal database since the true frequency of losses below this threshold is not zero. Furthermore, the major weaknesses of internal databases include the fact that internal loss data is not sufficient due to the short time period for data collection, and that most of the operational loss events are High Frequency Low Impact (HFLI) events. Since very few of the loss events are Low Frequency High Impact (LFHI), this leads to the inability of modeling the tail of operational risk distribution.

Internal loss data is not sufficient, and banks must also rely on relevant external loss data which provide insights into losses incurred by the entire industry. External loss data include information on operational loss events experienced by other firms and must be used to the enhancement of internal loss data for banks which are exposed to infrequent severe losses. Pooled industry and public data constitute external loss data.

The first type of providers of pooled industry data are the consortium of financial institutions which aim at creating a secure and confidential platform for sharing high quality individual loss data (internal loss data). Banking consortium databases are:

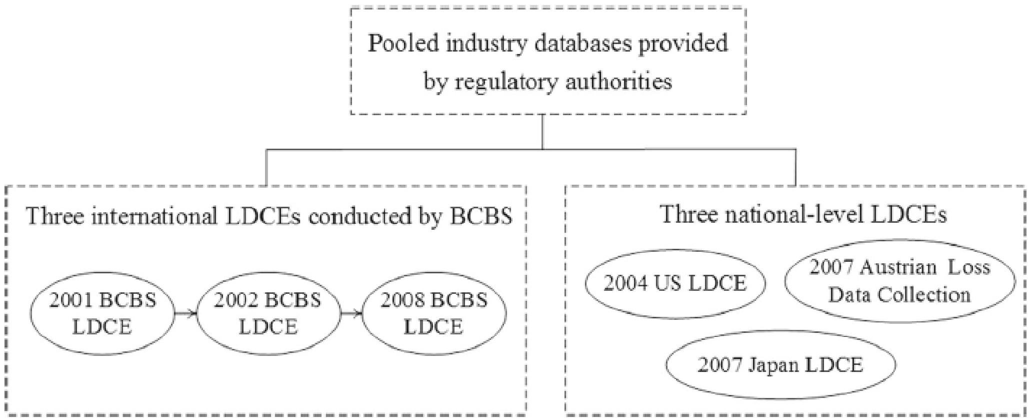
1. Operational Riskdata eXchange Global Banking Database (ORX-GBD)
2. Database Italiano delle Perdite Operative (DIPO) database
3. Daten Konsortium operationelle Risiken (DakOR) database
4. Deutsche Sparkassen-und Giroverband (DSGV) database
5. Global Operational Loss Database (GOLD)
6. Operational Loss Data Sharing Consortium Database (OLDSCD)
7. Credit Operational Risk Data Exchange (Cordex) database
8. Hungarian Operational Risk (HunOR) database

Moreover, insurance consortium databases are:

1. ORX Global Insurance Database (ORX-GID)
2. Operational Risk Insurance Consortium (ORIC) database.

While the second type of provides is the regulatory authorities which aim at constructing pooled industry databases for regulatory purposes. Regulatory authorities have suggested detailed data collection standards to compare loss experience across banks. The BCBS provides a forum for cooperation on banking supervisory matters as it is the primary global standard-setter for the regulation of banks. BCBS has conducted three international LDCEs in 2001, 2002, and 2008. Moreover, LDCEs conducted at the national level include the 2004 US LDCE and the 2007 Japan

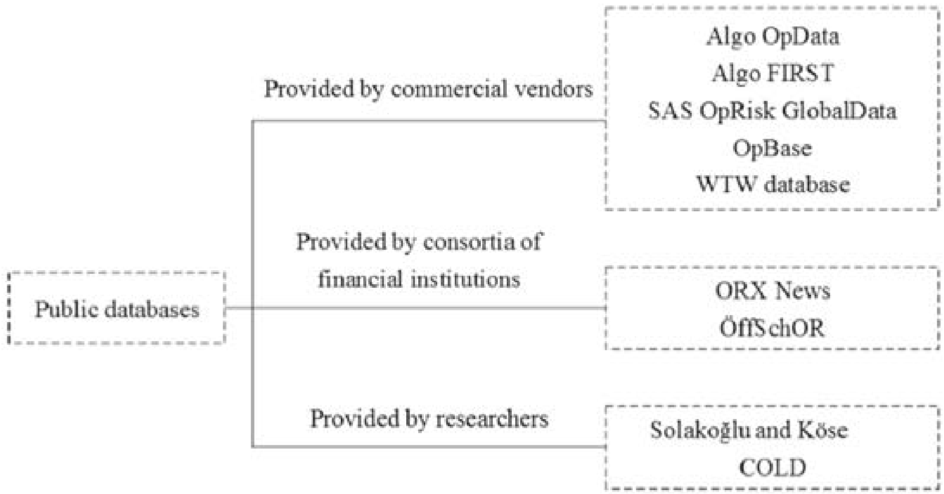
LDCE. Austria has also set up a confidential database for the sole use of the supervisory authority. These are summarized Figure 4.



	Participating banks	Total events	Threshold	Time period
2001 BCBS LDCE	30 Banks from 11 countries	27,000	€10,000	1998–2000
2002 BCBS LDCE	89 Banks from 19 countries	47,296	€10,000	2001
2008 BCBS LDCE	121 Banks from 17 countries	10.6 million	No-specified	No-specified
2004 US LDCE	23 US banks	1,526,765	No-specified	No-specified
2007 Japan LDCE	14 Japanese banks	156,112	No-specified	No-specified
2007 Austrian loss data collection	167 Austrian banks	42,000 (as of 2012)	No-specified	Since 2007

Figure 4: Industry Databases by Regulatory Authorities (Wei et al., 2018)

Public data is the second source of external loss data. It complements pooled industry data since the latter has a key limitation, which is it does not contain the data of non-participating banks. Public databases can take three forms according to different types of providers. Furthermore, amongst public databases the majority are commercial, provided by commercial vendors, which contain operational loss events in numerous financial institutions across the world over the past decades. Information on public databases are summarized in Figure 5 below.



Commercial database	Provider	Founded time	Threshold	Time window	Number of loss events
Algo FIRST	IBM Algorithmics	1998	$1 million	Since 1920	Over 14,500 (as of 2016)
Algo OpData	IBM Algorithmics	–	$1 million	Since 1972	Over 14,000 (as of 2017)
SAS OpRisk global data	SAS	2005	$1 million	Since 1831	32,000 (as of 2015)
WTW database	Willis Towers Watson	–	–	Since 1970	1413
OpBase	Aon Limited	–	–	–	–

Figure 5: Public and Commercial Databases (Wei et al., 2018)

Secondly, public databases provided by consortia of financial institutions include ORIC (a dataset with 15,000 publicly reported risk events from the financial services sector), ORX News (a database with over 5300 loss events in the global financial services industry from 2008 onward) and ÖffSchOR (it provides a detailed description of 2190 loss events across the world ranging from 1980 to 2015). Lastly, public databases provided by researchers include Solako˘glu and Köse (a self-collected database with 22 operational loss events in the Turkish banking sector between 1998 and 2007) and COLD (it contains 2132 operational loss events that occurred in the entire Chinese banking system over the years 1994–2012). All findings related to external loss databases are summarized in Figure 6 below.

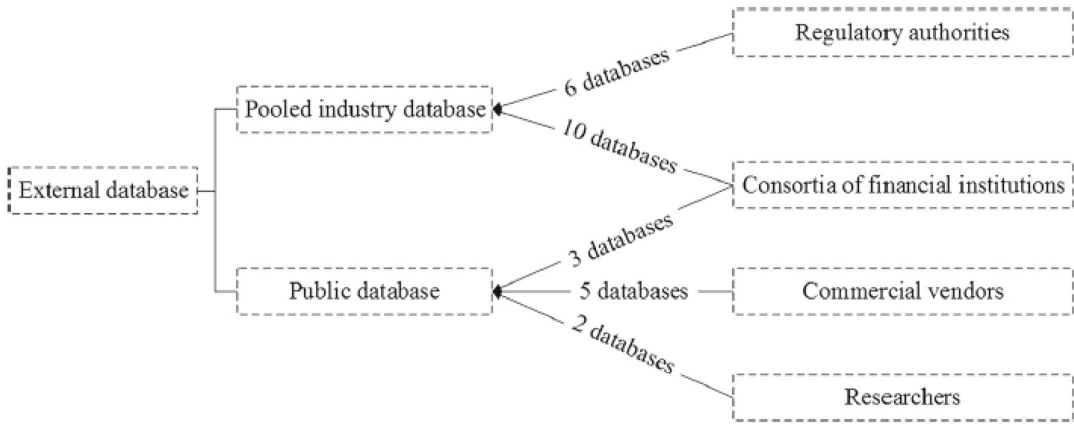


Figure 6: External Databases (Wei et al., 2018)

In scenario analysis, the knowledge of experienced business managers and risk management experts is used to develop rational evaluations of potential operational risk exposures. Scenario data have two quantitative components which are severity and frequency, and one descriptive component which is a detailed description of the scenario. The main reason for constructing a scenario database is to capture LFHI events which may not have happened in a bank's loss history. Individual banks and researchers gather and generate scenario data, which are then collected by consortia of financial institutions and regulatory authorities to establish scenario databases, which are needed in analyzing scenario data across the banking industry. A comprehensive review shows that there are a total of 7 scenario databases, including 1 scenario database provided by researchers, 4 scenario databases provided by consortia of financial institutions and 2 databases provided by regulatory authorities, which are summarized in Figure 7.

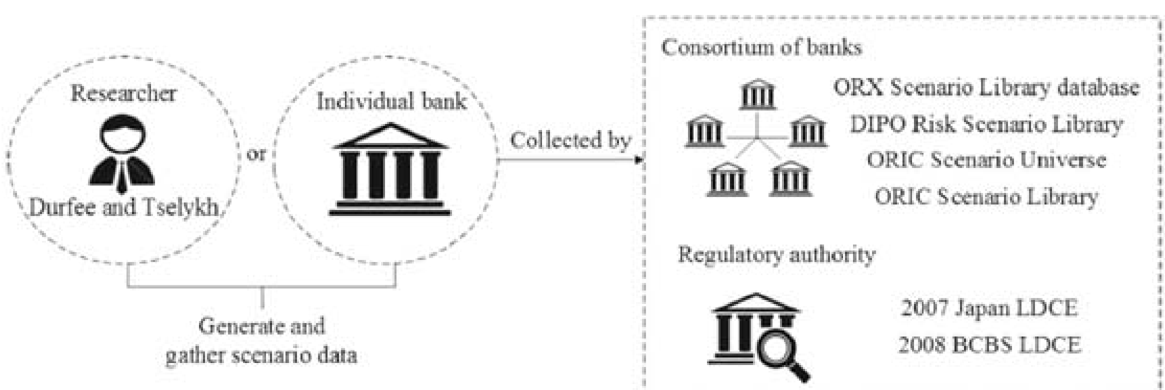


Figure 7: Breakdown of Scenario Databases (Wei et al., 2018)

The fourth and final data element of the AMA is the business environment and internal control factors (BEICFs). BEICFs provide forward-looking assessments of key business environment factors such as the rate of growth, employee turnover, and new product introductions, and internal control factors such as findings from the challenge process, internal audit results, and system downtime. The incorporation of these factors can recognize both improvements and deterioration in operational risk profiles in a more instant manner, which makes a bank's operational risk capital estimation sensitive to its changing operational risk profile. The 2008 BCBS LDCE has collected information of BEICFs only. The most commonly used BEICFs tools are risk and control self-assessments (RCSAs), audit results, key risk indicators (KRIs), key performance indicators (KPIs) and Key Control Indicators (KCIs). It is important to note that a KRI Library with about 2500 detailed KRI specifications has been established by the consortium of ORIC to help banks identify appropriate KRIs. Incorporating BEICFs into operational risk measurement is in its developmental stages particularly due to the subjectivity and structure of BEICF tools. Finally, the authors highlight that under the recently proposed Standardized Measurement Approach (SMA) of the BCBS, only the first AMA data element (internal loss) is used. The other three (external loss, scenario analysis, and BEICFs) are not included. Thus, the authors consider that the SMA may not provide a reliable estimation of operational risk because the use of each AMA data element is crucial for measuring operational risk.

Zhu et al., (2019) highlight the importance of the AMA in operational risk results within the Chinese banking industry and reject the decision of the Basel Committee to discard the AMA. The committee considers that this approach is complex and lacks comparability. Furthermore, it notes that the AMA was established with a significant degree of flexibility which was expected to be reduced over time, leading to the emergence of a best practice. The authors examine the convergence of operational risk results using different AMAs in the Chinese banking industry since 2006. Chinese researchers have conducted studies to improve the accuracy of the LDA under the AMA, noting that out of the 253 papers published on operational risk, 79 are written by Chinese researchers. Research has been conducted in better fitting loss frequency and severity distributions, using piecewise distributions, and introducing nonparametric methods. Moreover, dependencies across different business line/event type cells has been considered, which can have a substantial impact on operational risk estimation. The authors have conducted a comprehensive review of operational risk results in the Chinese banking industry and note that results varied considerably from less than 300 to more than 1500 (in billion CNY) in early years which involved the use of relatively small datasets and basic approaches. Yet, due to the improvements in the accuracy of the LDA based on advanced methods and approaches developed by Chinese researchers, result fluctuations clearly decline in later years. Thus, the AMA results narrow considerably over time. The authors conclude that as the AMAs have become more refined over the past decade, operational risk results for the Chinese banking industry have tended to converge. Thus, BCBS' decision to withdraw the AMA is not reasonable.

Nuugulu and Kock, (2020) propose robust operational risk capital allocation methods under the AMA through analytical work which attempts to fit several theoretical distributions to real banking and financial data. This work focuses only on modelling the severity of the loss data since it includes many difficulties. Two data sets are utilized, one from a Spanish Savings bank and another from a South African retail bank. The QQ-plot is used as a graphical goodness-of-fit technique, while the Kolmogorov-Smirnov (KS), the Anderson-Darling (AD), the Cramer-von Misses (CVM), and the Kuiper tests are used as empirical goodness-of-fit techniques to select the best distribution for each operational loss data set. The LD, WD, and PD are considered. Starting with the Spanish data, results show that the LD provides a fairly good fit based on the results of the Cramer-von Misses test. An estimation of the Value at Risk is 557,100 € by the LD and PD at 99.8% and 99.7% confidence levels respectively. Using the WD, the same VaR estimate is found at a 99.9% confidence level. Thus, 557,100 € is an accurate estimate of the capital requirement for operational risk of the Spanish Savings bank. Moving on to the South African retail bank, results show that the LD is a good fit for the data, however, there are some data points not corresponding with the fitted line. Therefore, a more heavier-tailed distribution is required. The WD fits the data better, yet there remains a significant number of deviating points. Finally, the PD provides a good fit and the best VaR estimates, but still a heavier distribution is required. At 99.8% confidence level, the VaR using the Pareto distribution is 9,669,000 R while the WD yields a VaR of 5,380,000 R at 99.9% confidence level. To this end, the POT method is used to estimate the VaR taking into account extreme loss events. For both data sets, a threshold is selected and the exceedances are fitted to the GPD. For the Spanish Savings bank data, the 97.4% VaR estimate is recorded as 407,100 €, thus the LD estimate is sufficient to cover operational losses on all fronts. While for the South African retail bank data, the 96.6% VaR estimate is recorded as 9,389,000 R, thus the PD estimate is accepted since the distribution best fits the data and it covers low severity losses and extreme events. The WD estimate is rejected as it does not cover operational losses on all fronts. The authors note that in both cases none of the best fit distributions could achieve the 99.9% VaR as proposed in Basel II. Yet, the authors highlight that both distributions provided superior fitness to the two datasets since their QQ-plots suggest evidence of under-fitting to some extent.

Chapter Three: Methodology

The Loss Distribution Approach

Modeling aggregate loss distributions is a central task in operational risk management. There are many approaches for calculating operational risk capital charge. The LDA is the most sophisticated one. Under the LDA, the bank or financial institution estimates the probability distributions of the severity and frequency using business line and risk category classes. It involves four main steps:

1. Loss frequency distribution
2. Loss severity distribution
3. Aggregate loss distribution by compounding (convolution)
4. Overall distribution (copula) and VaR measure

The application of these steps will yield the operational risk capital charge (Frachot et al., 2001).

3.1 Loss Frequency Distributions

In modeling the frequency of losses, the main task lies in determining a discrete random variable which represents the number of operational risk events observed. These events will occur with some probability p. There are numerous frequency distributions such as the binomial, negative binomial, geometric, … However, a detailed examination of the literature shows that the Poisson distribution is most commonly used in modeling frequency of loss events in operational risk.

The Poisson Distribution

The Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known constant mean rate and independently of the time since the last event. The Poisson distribution can also be used for the number of events in other specified intervals such as distance, area or volume.

A discrete random variable X is said to have a Poisson distribution, with parameter $\lambda > 0$.

$$f(k;\lambda) = \Pr(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

The positive real number λ is equal to the expected value of X and also to its variance.

$$\lambda = E(X) = Var(X)$$

3.2 Loss Severity Distributions

Any statistically based model of operational risk first requires fitting probability distributions to loss events on the severity of these losses arising from operational risks. The observed data may be actual data collected from a bank, financial institution, or insurance company. As well, it might be simulated data. The first step lies in considering a number of relevant distributions as examined in the literature. Then, distributional parameters are estimated and the distribution best fitting the data is selected. In this sense, the five most commonly used distributions for severity are presented.

The Normal Distribution

A normal distribution is a type of continuous probability distribution for a real-valued random variable. The general form of its probability density function is:

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

The parameter μ is the mean or expectation of the distribution (and also its median and mode), while the parameter σ is its standard deviation. The variance of the distribution is σ^2.

The Lognormal Distribution

A lognormal distribution is a continuous probability distribution of a random variable whose logarithm is normally distributed. Thus, if the random variable X is log-normally distributed, then Y = ln(X) has a normal distribution. A random variable which is log-normally distributed takes only positive real values.

A positive random variable X is log-normally distributed (X ~ Lognormal (μ_X^2, σ_X^2)) if the natural logarithm of X is normally distributed with mean μ and variance σ^2.

$$\ln(X) \sim N(\mu, \sigma^2)$$

Let Φ and φ be respectively the cumulative probability distribution function and the probability density function of the N (0, 1) distribution.

$$f_X(x) = \frac{d}{dx}\Pr(X \leq x) = \frac{d}{dx}\Pr(\ln X \leq \ln x) = \frac{d}{dx}\Phi\left(\frac{\ln x-\mu}{\sigma}\right)$$

$$f_X(x) = \varphi\left(\frac{\ln x-\mu}{\sigma}\right)\frac{d}{dx}\left(\frac{\ln x-\mu}{\sigma}\right) = \varphi\left(\frac{\ln x-\mu}{\sigma}\right)\frac{1}{\sigma x}$$

$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}}\exp\left(-\frac{(\ln x-\mu)^2}{2\sigma^2}\right)$$

The Weibull Distribution

The Weibull distribution is a continuous probability distribution. The probability density function of a Weibull random variable is:

$$f(x; \lambda, k) = \begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^{k}} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Where $k > 0$ is the shape parameter and $\lambda > 0$ is the scale parameter of the distribution

The Gamma Distribution

The gamma distribution is a two-parameter family of continuous probability distributions.

There are two different parameterizations in common use:

- With a shape parameter k and a scale parameter θ.
- With a shape parameter $\alpha = k$ and an inverse scale parameter $\beta = 1/\theta$, which is called a rate parameter.

In each of these forms, both parameters are positive real numbers.

Characterization using shape α and rate β:

A random variable X that is gamma-distributed with shape α and rate β is denoted:

$$X \sim \Gamma(\alpha, \beta) = \mathrm{Gamma}(\alpha, \beta)$$

$$f(x; \alpha, \beta) = \frac{\beta^{\alpha} x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} \text{ for } x > 0 \text{ and } \alpha, \beta > 0$$

The Log Logistic Distribution

The log-logistic distribution is a continuous probability distribution for a non-negative random variable. The log-logistic distribution is the probability distribution of a random variable whose logarithm has a logistic distribution. It is similar in shape to the log-normal distribution but has heavier tails.

$$f(x; \alpha, \beta) = \frac{(\beta/\alpha)(x/\alpha)^{\beta-1}}{\left(1 + (x/\alpha)^{\beta}\right)^{2}}$$

The parameter $\alpha > 0$ is a scale parameter and is also the median of the distribution. The parameter $\beta > 0$ is a shape parameter.

The above mentioned parametric approaches are used to estimate the severity distribution which fit high-frequency and low-severity loss events. However, operational risk loss distribution has a fat-tail characteristic since the loss is minor in most cases yet major in some extreme cases. Thus, loss events due to operational risk are divided to two categories which are the high-frequency low-severity and low-frequency high-severity loss events. (Wang et al., 2017). A thorough review of the literature shows that the Generalized Pareto Distribution is most commonly used to fit the tail of the severity distribution, which includes the low-frequency and high-severity losses.

The Generalized Pareto Distribution

The generalized Pareto distribution (GPD) is a family of continuous probability distributions. It is often used to model the tails of another distribution.

It is specified by three parameters: location μ, scale σ, and shape ξ.

$$f_{\xi}(z) = \begin{cases} (1+\xi z)^{\frac{-\xi+1}{\xi}} & \text{for } \xi \neq 0 \\ e^{-z} & \text{for } \xi = 0 \end{cases}$$

Where $z = \frac{x - \mu}{\sigma}$

3.3 Distribution Fitting

Distribution fitting comprises finding the appropriate probability distribution to a series of data concerning the repeated measurement of a variable phenomenon. Thus, the aim is to predict the probability or to forecast the frequency of occurrence of the magnitude of the phenomenon in a certain interval. The distribution giving a close fit is supposed to lead to good predictions.

There are many techniques of distribution fitting. In the parametric methods, the parameters of the distribution are calculated from the data series. The most prominent parametric method is the Maximum likelihood estimation (MLE) which involves maximizing a likelihood function, so that under the assumed statistical model the observed data is most probable. The point in the parameter space that maximizes the likelihood function is called the maximum likelihood estimate. The logic of maximum likelihood is both intuitive and flexible, and as such, the method has become a dominant means of statistical inference.

From a statistical standpoint, a given set of observations is a random sample from an unknown population. The goal of maximum likelihood estimation is to make inferences about the population that is most likely to have generated the sample, specifically the joint probability distribution of the random variables $\{y_1, y_2, \ldots\}$ not necessarily independent and identically distributed. Associated with each probability distribution is a unique vector $\theta = [\theta_1, \theta_2, \ldots, \theta_k]^T$ of parameters that index the probability distribution within a parametric family $\{f(. ; \theta) \mid \theta \in \Theta\}$, where Θ is called the parameter space, a finite-dimensional subset of Euclidean space.

Evaluating the joint density at the observed data sample $y = (y_1, y_2, \ldots, y_n)$ gives a real-valued function.

$$L_n(\theta) = L_n(\theta;y) = f_n(y;\theta)$$

This is called the likelihood function, and for independent and identically distributed random variables, $f_n(y;\theta)$ will be the product of univariate density functions.

The goal of maximum likelihood estimation is to find the values of the model parameters that maximize the likelihood function over the parameter space, that is

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \widehat{L}_n(\theta; y)$$

Intuitively, this selects the parameter values that make the observed data most probable.

The specific value $\hat{\theta} = \hat{\theta}_n(y) \in \Theta$ that maximizes the likelihood function L_n is called the maximum likelihood estimate. In practice, it is often convenient to work with the natural logarithm of the likelihood function, called the log-likelihood:

$$l(\theta; y) = \ln L_n(\theta; y)$$

While the domain of the likelihood function, the parameter space, is generally a finite-dimensional subset of Euclidean space, additional restrictions sometimes need to be incorporated into the estimation process. The parameter space can be expressed as:

$$\Theta = \{\theta: \theta \in R^k, h(\theta) = 0\}$$

Where $h(\theta) = [h_1(\theta), h_2(\theta), \ldots, h_r(\theta)]$ is a vector-valued function mapping R^k into R^r.

Estimating the true parameter θ belonging to Θ then, as a practical matter, means to find the maximum of the likelihood function subject to the constraint $h(\theta) = 0$.

In practice, restrictions are usually imposed using the method of Lagrange which, given the constraints as defined above, leads to the restricted likelihood equations

$$\frac{\partial l}{\partial \theta} - \frac{\partial h(\theta)^T}{\partial \theta} \lambda = 0 \text{ and } h(\theta) = 0$$

3.4 Goodness of Fit and Graphical Tests

The goodness-of-fit test is a statistical hypothesis test used to examine how well a sample data fits a certain distribution. Measures of goodness of fit typically summarize the discrepancy between observed values and the values expected under the model in question. There are multiple methods for determining goodness-of-fit. Some of the most popular methods used in statistics include the Kolmogorov-Smirnov test, the chi-square, the Anderson-Darling test, and the Shipiro-Wilk test.

There are many graphical and statistical tests for assessing the fit of a hypothesized severity of a loss probability distribution. The Q-Q and P-P plot are the most common plots used to assess the fitness of a distribution while the Kolmogorov-Smirnov test is the most common test used for goodness of fit.

3.4.1 Q-Q Plot

A Q-Q (quantile-quantile) plot is a probability plot, which is a graphical method for comparing both probability distributions, fitted and actual, by plotting their quantiles against each other. First, the set of intervals for the quantiles is chosen. A point (x, y) on the plot corresponds to one of the quantiles of the second distribution (y-coordinate) plotted against the same quantile of the first distribution (x-coordinate). Thus the line is a parametric curve with parameters that present the number of the interval for the quantile. If the two distributions being compared are similar, the points in the Q-Q plot will approximately lie on the line $y = x$. If the distributions are linearly related, the points in the Q-Q plot will approximately lie on a line, but not necessarily on the line $y = x$. Q-Q plots can also be used as a graphical means of estimating parameters in a location-scale family of distributions such as the gamma distribution.

3.4.2 P-P Plot

A P-P (probability-probability plot or percent-percent plot or P value plot) is a probability plot for assessing how closely two data sets agree; it plots the two cumulative distribution functions against each other. P-P plots are vastly used to evaluate the skewness of a distribution. The comparison line is the 45° line from (0,0) to (1,1) and the distributions are equal if and only if the plot falls on this line; any deviation indicates a difference between the distributions.

3.4.3 The Kolmogorov-Smirnov Test

The Kolmogorov–Smirnov test (or KS test) is a nonparametric test of the equality of continuous (or discontinuous) one-dimensional probability distributions that can be used to compare a sample with a reference probability distribution (one-sample KS test), or to compare two samples (two-sample KS test). It is named after Andrey Kolmogorov and Nikolai Smirnov.

The Kolmogorov-Smirnov test is defined by:

- H_o The data follow a specified distribution
- H_a The data do not follow the specified distribution

The Kolmogorov-Smirnov statistic quantifies the distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution, or between the empirical distribution functions of two samples. The null distribution of this statistic is calculated under the null hypothesis that the sample is drawn from the reference distribution (in the one-sample case) or that the samples are drawn from the same distribution (in the two-sample case). The empirical distribution function F_n for n independent and identically distributed (i.i.d.) ordered observations X_i is defined as

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

Where $I_{[-\infty,\, x]}(X_i)$ is the indicator function, equal to 1 if $X_I \leq x$ and equal to 0 otherwise.

The Kolmogorov-Smirnov statistic for a given cumulative distribution function F(x) is:

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

Where $\sup_x$ is the supremum of the set of distances. Intuitively, the statistic takes the largest absolute difference between the two distribution functions across all x values.

The p-value returned by the KS test has the same interpretation as other p-values. We reject the null hypothesis that the two samples were drawn from the same distribution if the p-value is less than the significance level. The smaller the p-value, the stronger the evidence that you should reject the null hypothesis.

Many times, the KS test results show that several distributions are a good fit for the data. In order to select the best fitted distribution, the Akaike Information Criterion (AIC) is considered. It is an estimator of prediction error and thereby relative quality of statistical models for a given set of data. AIC deals with both the risk of over fitting and the risk of under fitting.

The application of the AIC involves starting with a set of candidate models, and then finding the models' corresponding AIC values. There will almost always be information lost due to using a candidate model to represent the "true model". The aim is to select, from among the candidate models, the model that minimizes the information loss.

3.5 Threshold Selection

Since operational risk loss distribution has a fat-tail characteristic, threshold selection is a key to parameter estimation and prediction of loss data. If a low threshold is chosen, then the risk of obtaining biased estimates is incurred. While if a high threshold is selected, then few data points will be obtained and thus the estimation will be subject to high standard errors. There are many threshold selection methods, and the most commonly adopted ones are presented below.

3.5.1 The Mean Excess Plot

The mean excess plot is a useful method to determine the threshold u.

Using the mean excess function $e(u) = E(X - u \mid X > u)$, we can obtain the following:

$$e(u) = \frac{\beta + \alpha u}{1 - \alpha}, \qquad \beta + \alpha u > 0$$

For a given sample $x_1, x_2, \ldots x_n$, its mean excess loss function e(u) is defined as:

$$e(u) = \sum_{i=1}^{n} \frac{(x_i - u)^+}{N_u}$$

Where N_u is the number of data points exceeding the threshold u:

$$(x_i - u)^+ = \begin{cases} x - u, & x > u \\ 0 & x \leq u \end{cases}$$

That is, the sum of excess over the threshold u divided by the number of data points exceeding the threshold u. An empirical plot that apparently follows a reasonably straight line with a positive gradient above a certain value of u indicates that the u can be chosen as a threshold.

3.5.2 The Hill Plot

The order of observations is $X_1 > X_2 > \ldots > X_k > \ldots > X_n$, and the Hill estimator based on the order statistics is:

$$\gamma_{k,\,n} = \frac{1}{k} \sum_{i=1}^{k} \log \left[\frac{X_i}{X_k}\right]$$

The Hill plot is defined as a scatter plot which is composed by $(k, \gamma^{-1}_{k,\,n})$. X_k is selected as a threshold and the subscript k is the abscissa of the Hill plot from where the scatter plot begins to stabilize.

3.6 Aggregate Loss Distribution

It is possible to generate sample values that represent aggregate operational risk losses given the severity and frequency of a loss probability model. Using the frequency and severity of loss data, we can simulate aggregate operational risk losses and then use these simulated losses for the calculation of the operational risk capital charge. The best way to achieve the aggregate loss distribution is to collect data on frequency and severity of losses for a particular operational risk type and then fit frequency and severity of loss models to the data. Thus, convolution is used to derive the distribution of a sum of two distributions. The aggregate loss distribution can be found by compounding the distributions for severity and frequency of operational losses over a fixed period such as a year. One of the most commonly used convolution methods for deriving the aggregate operational risk distribution is the Monte Carlo method, and it is presented below.

The Monte Carlo Method

It involves the following steps:

1. Choose a severity of loss and frequency of loss probability model.
2. Generate n number of losses daily or weekly regarding the frequency of loss distribution.
3. Generate n losses X_i ($i = 1, \dots, n$) regarding the loss severity distribution.
4. Repeat steps 2 and 3 for N= 365 (for daily losses) or N = 52 (for weekly losses). Summing all the generated X_i to obtain S which is the annual loss.
5. Repeat the steps 2 to 4, at least 5000 times, to obtain the annual aggregate loss distribution.

3.7 Overall Loss Distribution

The Basel requirements for the calculation of the total operational capital charge involves a simple summation of the capital charges of all 56 risk classes. In this sense, a perfect positive dependence between the risks is implicitly assumed. However, this will yield large capital requirements and therefore it is crucial to investigate the dependence structure. Conventionally, correlation is used to model dependence between variables, but latest research works show the superiority of copula over correlation for modeling dependence due to higher flexibility of the copula and its ability to model dependence between extreme events.

A copula is a multivariate joint distribution defined on the n-dimensional unit cube $[0, 1]^n$ in a way that every marginal distribution is uniform in the interval [0, 1]. There are various copula functions, however according to the literature, a multidimensional t-copula is commonly used in modeling operational risk. The t-copula exhibits tail dependence, which is appealing in operational risk modeling and it is capable of modeling dependence in the tail without giving up the flexibility to model dependence in the center (Momen et al., 2012).

A Multivariate t-copula (MTC) is defined as follows:

$$T_{R,v}(u_1, u_2, \ldots, u_n) = t_{R,v}(t_v^{-1}(u_1), t_v^{-1}(u_2), \ldots, t_v^{-1}(u_n))$$

Where R is a symmetric, positive definite matrix with diag(R) = $(1, 1, \ldots, 1)^T$ and $t_{R,v}$ is the standardized multivariate Student's t distribution with correlation matrix R and v degrees of freedom. $t_v - 1$ is the inverse of the univariate cdf of Student's t distribution with v degrees of freedom. Using the canonical representation, it turns out that the copula density for the MTC is:

$$C_{R,v}(u_1, u_2, \ldots, u_n) = |R|^{-\frac{1}{2}} \frac{\Gamma(\frac{v+n}{2})}{\Gamma(\frac{v}{2})} \left(\frac{\Gamma(\frac{v}{2})}{\Gamma(\frac{v+1}{2})} \right)^n \frac{(1+\frac{1}{v}\varsigma^T R^{-1}\varsigma)^{-\frac{v+n}{2}}}{\prod_{j=1}^{n}(1+\frac{\varsigma_j^2}{v})^{-\frac{v+1}{2}}}$$

Where $\varsigma_j = t_v^{-1}(u_j)$, Using the t-copula, the capital charge can be calculated (Momen et al., 2012).

3.8 Capital Charge

Under the LDA, the capital charge or known as the Capital-at-Risk is a Value-at-Risk measure of risk. Given some confidence level $\alpha \in (0, 1)$ the Value-at- Risk (VaR) at the confidence level α is given by the smallest number *l* in a way that the probability that the loss L exceeds *l* is not larger than $(1 - \alpha)$. The capital charge is the summation of the VaR at confidence level α of each of the different risk classes required by the BCBS. VaR can be defined as "a statistical estimation of a portfolio loss with the property that, with a given (small) probability, we stand to incur that loss or more over a given (typically short) holding period" (Embrechts et al., 2003).

$$VAR_a = \inf\{l \in R: P(L > l) \leq 1 - a\}$$

$$VAR_a = \inf\{ l \in R: F_L \geq a\}$$

The right equality assumes an underlying probability distribution, which makes it true only for the parametric VaR. The left equality means that we are $100(1 - \alpha)\%$ confident that the loss in the related period will not be larger than the VaR (Momen et al., 2012).

The Basel framework does not specify the analytical approaches for operational risk, but a bank must show that its adopted measure for operational risk meets a soundness standard such as a one year holding period and a 99.9^{th} percentile confidence interval.

Chapter Four: Data, Results, and Discussion

4.1 Operational Risk Data

There are numerous operational risk databases as indicated in Chapter 2. However, these databases are not open to individuals, only corporations or institutions can access these databases. In addition access is granted on condition that the joining member has to share their own operational risk data, and that is to ensure the sustainability of the database. In this sense, researchers in this field suffer tremendously in search for data to conduct studies. Even journal papers do not contain the full dataset required. In the case of this study, thorough research was conducted to obtain a complete set of operational risk data from any database or researcher in the field. However, all the attempts were unsuccessful in obtaining real operational risk data. Thus, the final resort in our case was to obtain simulated data online. The dataset utilized in this research is obtained from GitHub. The data obtained consists of 3,192 operational loss events from the year 2009 and until 2018. For each loss event, the following terms are provided, namely ID, Date, Business Line, Risk Category, Gross Loss Amount, and Recovery Amount. The major benefit of this dataset is that the operational risk loss events are categorized as per the Basel requirements for business line and risk category.

This makes the application of the LDA under the Basel requirements possible. Under Basel II, the LDA requires to quantify distributions for frequency and severity of operational risk loss data for each cell of the 56-cell matrix combining the eight business lines and seven event types highlighted in Chapter 2 within Basel II Capital Accord and presented in Table 13. The eight business lines (BL) are listed by row and the seven risk categories (RC) are listed in by columns.

In the case of this research, the simulated data obtained is divided over eight business lines and six risk categories. Risk categories are equivalent to event types. The 7th risk category which does not have any loss events assigned to it is “Business Disruption & System Failures”. Therefore, in this case, the data is spread over a 48-cell matrix, that is 8 business lines *x* 6 risk categories.

The implementation of the LDA is accomplished using R programming language: R studio 4.0.3.

Table 13: The 56-cell Matrix

BL \ RC	Internal Fraud	External Fraud	Employment Practices & Workplace Safety	Clients, Products, & Business Practices	Damage to Physical Assets	Business Disruption & System Failures	Execution, Delivery, & Process Management
Corporate Finance							
Trading & Sales							
Retail Banking							
Commercial Banking							
Payment & Settlements							
Agency Services							
Asset Management							
Retail Brokerage							

Tables 14 and 15 show the abbreviations used for business lines and risk categories in this study.

Table 14: Business Lines' Abbreviations

Business Lines	
AG	Agency Services
AM	Asset Management
CB	Commercial Banking
CF	Corporate Finance
PS	Payment & Settlements
RBA	Retail Banking
RBR	Retail Brokerage
TS	Trading & Sales

Table 15: Risk Categories' Abbreviations

Risk Categories	
CPBP	Clients, Products, & Business Practices
DPA	Damage to Physical Assets
EDPM	Execution, Delivery, & Process Management
EF	External Fraud
EPWS	Employment Practices & Workplace Safety
IF	Internal Fraud

Table 16 presents the count and sum of loss amounts per year, while Tables 17 and 18 show a breakdown of these parameters over the 48-cell matrix respectively.

Table 16: Count and Sum of Loss Amounts over All Years

Year	Count of Loss Amount	Sum of Loss Amount
2009	237	41,141,849.1
2010	210	29,264,198.6
2011	475	25,408,481.86
2012	572	47,028,002.82
2013	191	26,751,725.55
2014	450	55,772,861.78
2015	196	36,805,607.01
2016	230	9,681,592.7
2017	335	29,799,515.61
2018	296	155,363,004
TOTAL	**3,192**	**457,016,838.98**

Table 17: Count of Loss Amount Breakdown

Count of Loss Amount							
RC / BL	CPBP	DPA	EDPM	EF	EPWS	IF	TOTAL
AG	610			136		54	**800**
AM			114			157	**271**
CB	318			51	124	50	**543**
CF		127	27	42			**196**
PS		251	58		78	80	**467**
RBA			15			60	**75**
RBR	421		133			63	**617**
TS		207		16			**223**
TOTAL	**1349**	**585**	**347**	**245**	**202**	**464**	**3192**

Table 18: Sum of Loss Amount Breakdown

Sum of Loss Amount							
RC / BL	CPBP	DPA	EDPM	EF	EPWS	IF	TOTAL
AG	73064225.45			7345959.62		3121818.72	**83532003.79**
AM			5764053.66			9041258.9	**14805312.56**
CB	30784949.9			2609418.7	16584413.53	19109419.08	**69088201.21**
CF		114999120.8	524084.59	35239168.42			**150762373.8**
PS		13486537.48	1880023.27		1897672.73	23204153.71	**40468387.19**
RBA			26975464.28			2055989.24	**29031453.52**
RBR	15616867.91		23860823.48			1363736.17	**40841427.56**
TS		28317105.84		170573.48			**28487679.32**
TOTAL	**119466043.3**	**156802764.1**	**59004449.28**	**45365120.22**	**18482086.26**	**57896375.82**	**457016838.98**

One of the main restrictions of this research work remains the dataset utilized. For this research, it is impossible to implement the LDA at the level of the business line *x* risk category cells, and that is due to the count of the loss events. Table 19 shows the breakdown of the loss events of 2009 over the 48-cell matrix addressed.

Table 19: Count of Loss Amount Breakdown for 2009

Count of Loss Amount – 2009							
BL \ RC	**CPBP**	**DPA**	**EDPM**	**EF**	**EPWS**	**IF**	**TOTAL**
AG	48			7		6	**61**
AM			8			15	**23**
CB	22			4	8	7	**41**
CF		11		5			**16**
PS		22	4		3	3	**32**
RBA			1			6	**7**
RBR	30		10			6	**46**
TS		10		1			**11**
TOTAL	**100**	**43**	**23**	**17**	**11**	**43**	**237**

Even before excluding the low loss amounts and dividing the data into body and tail, which will be addressed next, the data is insufficient for modeling as a cell matrix under the LDA. As can be seen, there are 26 cells without any loss events to start with, and 14 cells with 10 or less loss events. This is the case for all the years; the same 26 cells do not have loss events and numerous other cells have 10 or less loss events. This situation is not so uncommon. A review of the literature shows that many research studies have been conducted with a certain number of business lines or risk categories only due to either data unavailability or very low loss event count. Therefore, in the case of this research, the LDA will be implemented per business line and per risk category and at the end, the capital requirement will determined for each case.

4.2 Results and Discussion

4.2.1 De Minimis Gross Loss Threshold

The Basel committee addresses certain standards which banks must meet in the loss data collection process. A bank's internal loss data must be comprehensive in that it captures all material activities and exposures from all appropriate sub-systems and geographic locations. A bank must have an appropriate *de minimis* gross loss threshold for internal loss data collection, for example €10,000. The data used in this research is currency agnostic and therefore gross loss amounts are considered to be incurred in dollars. Therefore, the *de minimis* gross loss threshold was set as $10,000 whereby all gross loss amounts below this threshold were categorized as "Low" and excluded all together.

4.2.2 Body and Tail of Loss Data

Since the operational risk loss data is characterized by a fat tail, a threshold must be selected as a truncation point. The threshold is determined to separate the body from the tail. Therefore, both the Mean Excess Plot and the Hill Plot methods were implemented to calculate the said threshold. Table 20 summarizes the threshold values for business lines and risk categories of 2009.

Similarly, these two methods are implemented all throughout the remaining years.

Table 20: Threshold Values for Business Lines and Risk Categories for 2009

Set	Mean Excess Plot	Hill Plot	Selection (Minimum)
AG	225,000	225,000	225,000
CB	300,000	300,000	300,000
RBR			No Tail
PS	250,000	250,000	250,000
AM	200,000	200,000	200,000
CF	200,000	200,000	200,000
RBA			No Tail
TS	275,000	275,000	275,000
CPBP	300,000	300,000	300,000
IF	250,000	250,000	250,000
DPA	550,000	825,000	550,000
EDPM			No Tail
EF			No Tail
EPWS	250,000	250,000	250,000

The results of the two adopted methods are the same or very close for most of the classes. Some differences are identified in rare instances. In all cases and as a safety factor, for each class, the minimum of the two thresholds is selected to divide the data. All gross loss amounts above the selected thresholds are categorized as "Tail", and the remaining losses which are below the said threshold, yet above the *de minimis* threshold, are categorized as "Body". Furthermore, it is important to note that certain risk classes did not yield any threshold value indicating that these classes do not include high-severity and low-frequency events.

4.2.3 Count of Loss Events

The data used in this research imposes many restrictions. Tables 21 and 22 summarize the count of loss events throughout the years for each business line and risk category respectively.

There are numerous risk classes, business lines or risk categories, which include a small number of loss events such as the case of RBA, or TS, or EPWS for example. It is important to note that these are the total number of loss events per risk class.

Table 21: Count of Loss Events for All Years for each Business Line

BL	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	Total
AG	61	56	126	139	49	119	45	51	77	77	**800**
AM	23	15	37	48	14	43	20	24	22	25	**271**
CB	41	40	81	98	32	85	28	30	56	52	**543**
CF	16	11	28	35	15	22	12	15	27	15	**196**
PS	32	29	64	84	21	61	28	44	56	48	**467**
RBA	7	3	14	13	8	9	4	4	10	3	**75**
RBR	46	41	96	117	37	80	36	48	70	46	**617**
TS	11	15	29	38	15	31	23	14	17	30	**223**
Total	**237**	**210**	**475**	**572**	**191**	**450**	**196**	**230**	**335**	**296**	**3192**

Table 22: Count of Loss Events for All Years for each Risk Category

RC	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	Total
CPBP	100	88	204	242	89	206	74	89	137	120	**1349**
DPA	43	38	78	103	38	73	46	41	67	58	**585**
EDPM	23	27	47	70	20	44	19	28	35	34	**347**
EF	17	17	43	39	12	31	16	18	26	26	**245**
EPWS	11	14	26	34	10	34	9	17	22	25	**202**
IF	43	26	77	84	22	62	32	37	48	33	**464**
Total	**237**	**210**	**475**	**572**	**191**	**450**	**196**	**230**	**335**	**296**	**3192**

Therefore, it is essential to examine the number of loss events which constitute the "Body" of each risk class. In this sense, the Table 23 summarizes these values for 2009.

Table 23: Count of Loss Events for the Body of each Risk Class for 2009

Set	Total	Body	Tail
2009 (Total)	237	111	7
AG	61	26	1
AM	23	10	1
CB	41	19	2
CF	16	9	1
PS	32	16	1
RBA	7	5	0
RBR	46	19	0
TS	11	6	2
CPBP	100	39	1
DPA	43	23	2
EDPM	23	13	0
EF	17	11	0
EPWS	11	4	1
IF	43	22	2

Distribution fitting on small size samples is statistically not significant. Examining the count under "Body" of each risk class dictates that certain classes must be merged. Therefore, in this case, four business lines, particularly AM, CF, RBA, and TS are merged into one class titled "other". While three risk categories, particularly EDPM, EF, and EPWS are merged into one class titled "rest". This is similarly implemented for all years with certain provisions where needed. For these merged classes, the threshold value is set to be the minimum threshold among constituents.

4.2.4 Frequency Distribution

The information provided for each operational risk event in the dataset used is limited to: ID, Date, Business Line, Risk Category, Gross Loss Amount, and Recovery Amount. Thus, this imposes a major restriction on modeling the frequency of the losses. In this sense, an alternative method is proposed and implemented which is the Poisson process. Since the date of each loss event is available, then the time between two consecutive events is assumed to be exponentially distributed. Thus, the collection of these points (operational risk events) forms a Poisson process, and therefore the frequency distribution will follow a Poisson distribution which is consistent with the literature.

A Poisson Process is a model for a series of discrete events where the average time between events is known, but the exact timing of events is random. The arrival of an event is independent of the event before (waiting time between events is memoryless). A Poisson process is defined in terms of the sequence of inter-arrival times $X_1, X_2 \ldots$ which are defined to be independent and identically distributed. A Poisson process is a renewal process in which the inter-arrival intervals have an exponential distribution function. For some real $\lambda > 0$, each X_i has the density $f_X(x) = \lambda \exp(-\lambda x)$ for $x \geq 0$. The parameter λ is called the rate of the process.

For a Poisson process of rate λ, and for any $t > 0$, the PMF for $N(t)$ which is the Poisson counting process, that is the number of arrivals in $(0, t]$, is given by the Poisson PMF:

$$p_{N(t)}n = \frac{(\lambda t)^n \exp(-\lambda t)}{n!}$$

For each risk class, the operational risk loss events were first sorted in chronological order and the Poisson process was implemented. The results of the process were the determination of the lambda (λ) parameter of the Poisson distribution for frequency modeling.

Table 24 below summarizes the λ values for business lines and risk categories of 2009. Similarly, this process is implemented all throughout the remaining years.

Table 24: Poisson's Lambda Values of Business Lines and Risk Categories for 2009

Set	Poisson's Lambda	Set	Poisson's Lambda
AG	0.98	**CPBP**	2.90
CB	0.45	**IF**	0.41
RBR	0.50	**DPA**	0.75
PS	0.32	**Rest**	1.11
Other	1.18		

4.2.5 Severity Distribution

Regarding the severity distribution, first a logarithmic transformation of the gross loss amount was implemented since the original data didn't fit any common distribution. Afterward, the five most commonly used distributions in operational risk modeling are implemented, namely normal, lognormal, gamma, weibull, and log logistic. For each risk class, business line or risk category, the "Body" of the data is fitted against these distributions. Accordingly, Q-Q and P-P plots are generated to graphically assess the fitness of said distributions and the KS test is implemented to determine the goodness of fit through the p-value and the AIC. In this study, a p-value < 0.05 is considered statistically significant. It indicates strong evidence against the null hypothesis, as there is less than a 5% probability the null is correct and the results are random. Furthermore, to select the best fitting distribution, the AIC value of distributions with p-value > 0.05 are compared, and the one with the minimum AIC is selected.

Furthermore, regarding the operational risk losses belonging to the "Tail", these losses are fitted against the Generalized Pareto distribution. However, as addressed in section 4.2.3, the main restriction in this research is in the number of operational risk events. For all risk classes, there are no enough observations to statistically fit the data with a distribution. Similarly, these steps are implemented for each year.

The results for the year 2009 are presented and both Q-Q and P-P plots for each risk class show that all five distributions are a good fit for the data. This is determined by examining the closeness of the data points to the best fit line. However, this graphical interpretation remains subjective and the KS test results must be studied to quantitatively select the best fit distribution.

Business Line Severity Distribution Fitting

1. Agency Services (AG)

Table 25 presents the total records for AG in 2009 as well the number of losses in body, tail, and low. In this case, fitting distribution in the tail is discarded since it contains only one observation.

Table 25: Total, Body, Tail, and Low Count of Losses of AG for 2009

Set	Total	Body	Tail	Low
2009_AG	61	26	1	34

Figure 8 presents the histogram of the frequency of gross loss amounts in AG for 2009.

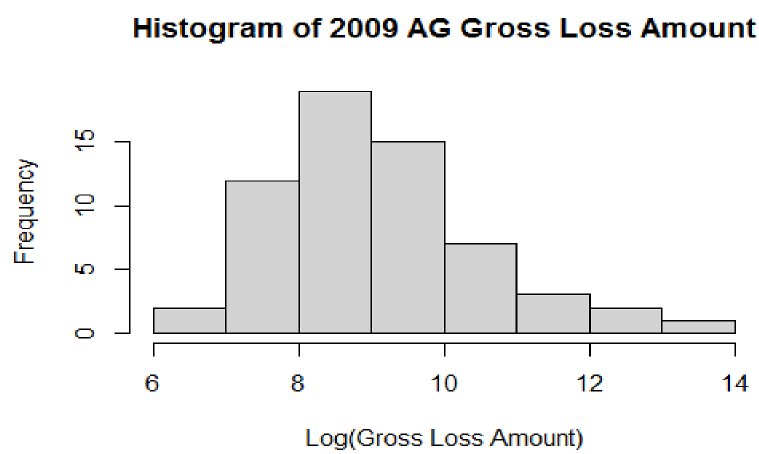


Figure 8: Histogram of the Frequency of Gross Loss Amounts in AG for 2009

Figure 9 presents the Q-Q and P-P plots respectively.

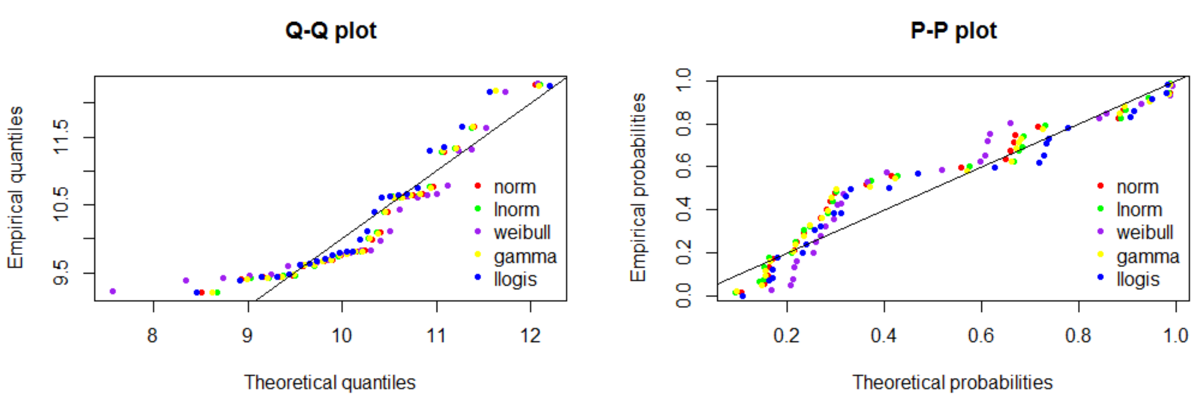


Figure 9: Q-Q and P-P Plots of the Fitted Severity Distributions of AG for 2009

An examination of the Q-Q and P-P plots shows that all five distributions are a good fit; that is based on their closeness to the best fit line.

Table 26 presents the estimated parameters of the fitted distributions in addition to the AIC and p-value of the KS test.

Table 26: Parameters of the Fitted Distributions, AIC, and p-value of AG for 2009

Distribution	Parameter 1	Parameter 2	AIC	p-value	Accepted
Normal	10.28	0.85	69.63	0.21	TRUE
Lognormal	2.33	0.08	67.80	0.23	TRUE
Weibull	11.38	10.69	76.00	0.31	TRUE
Gamma	151.04	14.70	68.38	0.22	TRUE
Log logistic	21.47	10.15	68.70	0.40	TRUE

The KS test results indicate that all distribution have a p-value > 0.05. Hence, the AIC values are examined to select the best fit distribution among the five. In this case, the lognormal distribution records the minimum AIC and therefore it is selected as the severity distribution.

2. Commercial Banking (CB)

Table 27 presents the total records for CB in 2009 as well the number of losses in body, tail, and low. In this case, fitting distribution in the tail is discarded since it contains only two observations.

Table 27: Total, Body, Tail, and Low Count of Losses of CB for 2009

Set	Total	Body	Tail	Low
2009_CB	41	19	2	20

Figure 10 presents the histogram of the frequency of gross loss amounts in CB for 2009.

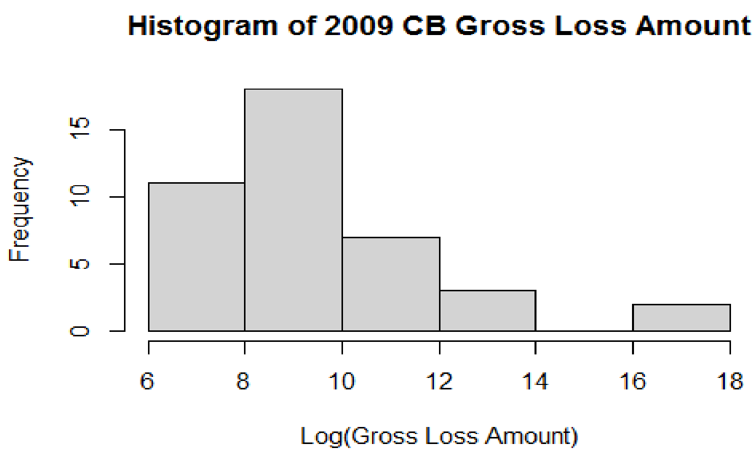


Figure 10: Histogram of the Frequency of Gross Loss Amounts in CB for 2009

Figure 11 presents the Q-Q and P-P plots respectively.

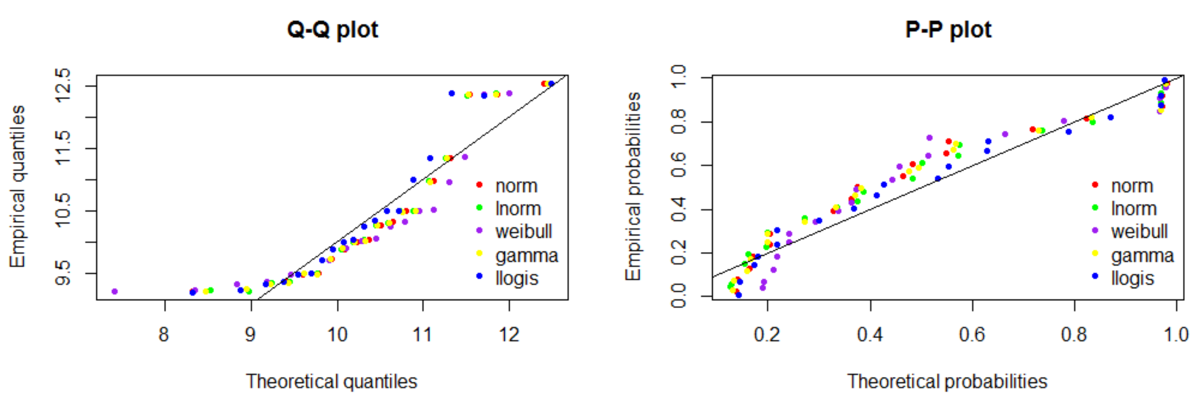


Figure 11: Q-Q and P-P Plots of the Fitted Severity Distributions of CB for 2009

An examination of the Q-Q and P-P plots shows that all five distributions are a good fit; that is based on their closeness to the best fit line.

Table 28 presents the estimated parameters of the fitted distributions in addition to the AIC and p-value of the KS test.

Table 28: Parameters of the Fitted Distributions, AIC, and p-value of CB for 2009

Distribution	Parameter 1	Parameter 2	AIC	p-value	Accepted
Normal	10.36	1.06	59.99	0.48	TRUE
Lognormal	2.33	0.10	58.35	0.64	TRUE
Weibull	9.46	10.87	64.13	0.27	TRUE
Gamma	101.55	9.80	58.86	0.58	TRUE
Log logistic	17.82	10.19	58.91	0.79	TRUE

The KS test results indicate that all distribution have a p-value > 0.05. Hence, the AIC values are examined to select the best fit distribution among the five. In this case, the lognormal distribution records the minimum AIC and therefore it is selected as the severity distribution.

3. Retail Brokerage (RBR)

Table 29 presents the total records for RBR in 2009 as well the number of losses in body, tail, and low. In this case, fitting distribution in the tail is discarded since it contains zero observations.

Table 29: Total, Body, Tail, and Low Count of Losses of RBR for 2009

Set	Total	Body	Tail	Low
2009_RBR	46	19	0	27

Figure 12 presents the histogram of the frequency of gross loss amounts in RBR for 2009.

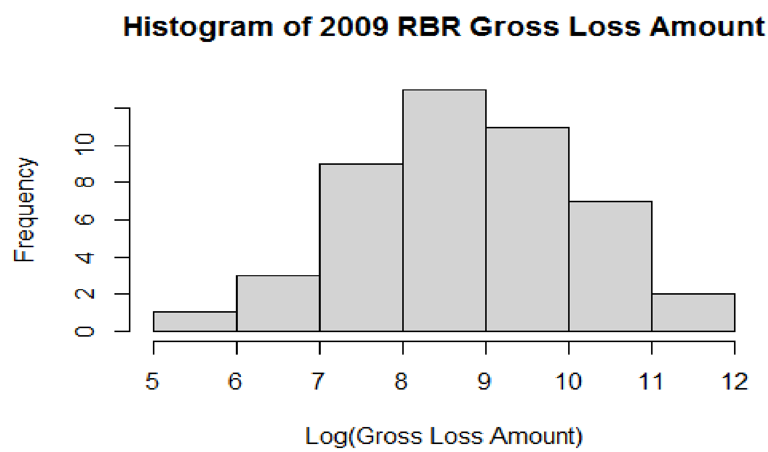


Figure 12: Histogram of the Frequency of Gross Loss Amounts in RBR for 2009

Figure 13 presents the Q-Q and P-P plots respectively.

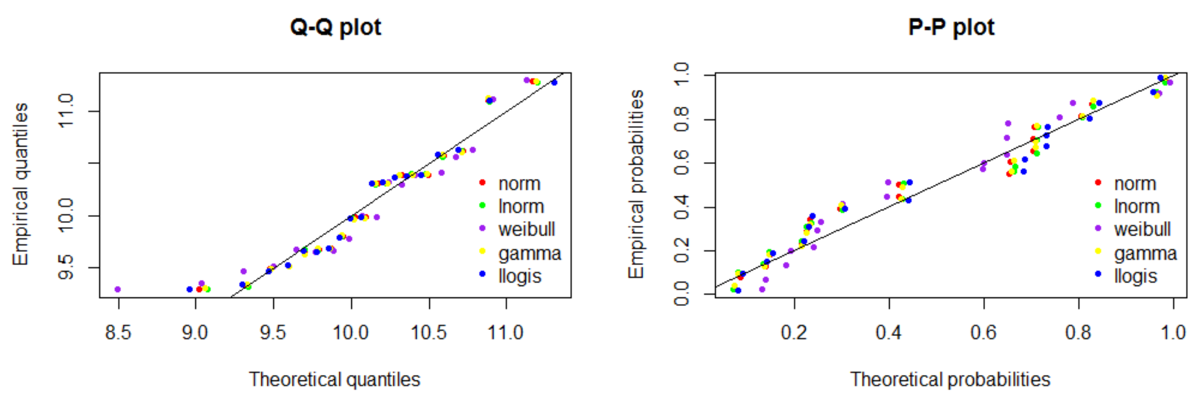


Figure 13: Q-Q and P-P Plots of the Fitted Severity Distributions of RBR for 2009

An examination of the Q-Q and P-P plots shows that all five distributions are a good fit; that is based on their closeness to the best fit line.

Table 30 presents the estimated parameters of the fitted distributions in addition to the AIC and p-value of the KS test.

Table 30: Parameters of the Fitted Distributions, AIC, and p-value of RBR for 2009

Distribution	Parameter 1	Parameter 2	AIC	p-value	Accepted
Normal	10.10	0.55	35.46	0.83	TRUE
Lognormal	2.31	0.05	35.03	0.83	TRUE
Weibull	18.18	10.37	38.91	0.80	TRUE
Gamma	336.94	33.37	35.16	0.83	TRUE
Log logistic	31.02	10.06	36.20	0.69	TRUE

The KS test results indicate that all distribution have a p-value > 0.05. Hence, the AIC values are examined to select the best fit distribution among the five. In this case, the lognormal distribution records the minimum AIC and therefore it is selected as the severity distribution.

4. Payment and Settlements (PS)

Table 31 presents the total records for PS in 2009 as well the number of losses in body, tail, and low. In this case, fitting distribution in the tail is discarded since it contains only one observation.

Table 31: Total, Body, Tail, and Low Count of Losses of PS for 2009

Set	Total	Body	Tail	Low
2009_PS	32	16	1	15

Figure 14 presents the histogram of the frequency of gross loss amounts in PS for 2009.

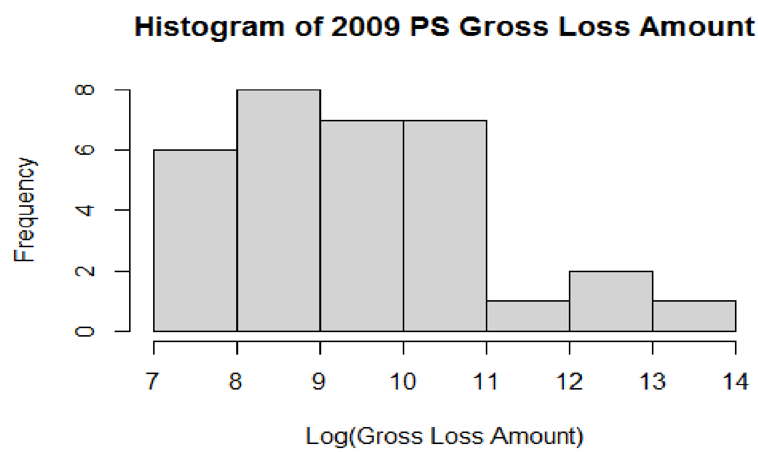


Figure 14: Histogram of the Frequency of Gross Loss Amounts in PS for 2009

Figure 15 presents the Q-Q and P-P plots respectively.

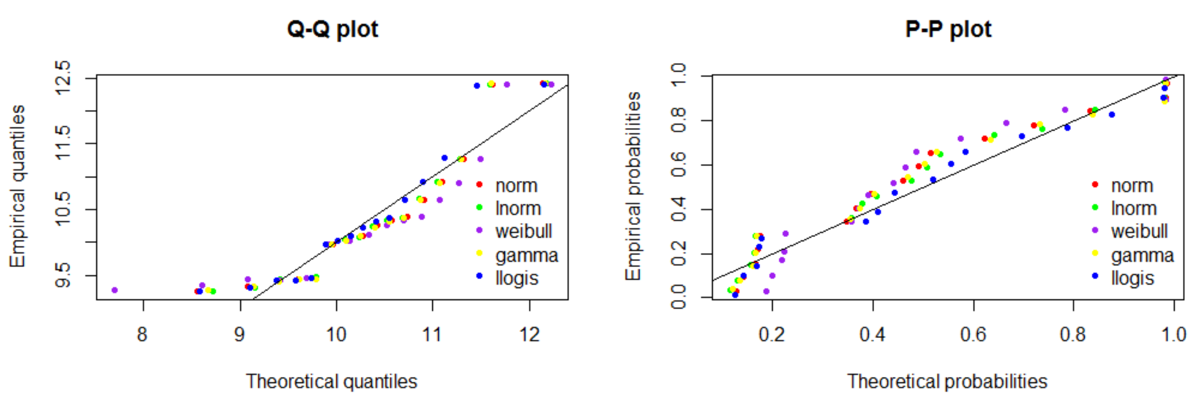


Figure 15: Q-Q and P-P Plots of the Fitted Severity Distributions of PS for 2009

An examination of the Q-Q and P-P plots shows that all five distributions are a good fit; that is based on their closeness to the best fit line.

Table 32 presents the estimated parameters of the fitted distributions in addition to the AIC and p-value of the KS test.

Table 32: Parameters of the Fitted Distributions, AIC, and p-value of PS for 2009

Distribution	Parameter 1	Parameter 2	AIC	p-value	Accepted
Normal	10.35	0.96	48.14	0.67	TRUE
Lognormal	2.33	0.09	46.85	0.79	TRUE
Weibull	10.15	10.81	52.07	0.47	TRUE
Gamma	121.90	11.78	47.25	0.75	TRUE
Log logistic	19.77	10.21	47.03	0.90	TRUE

The KS test results indicate that all distribution have a p-value > 0.05. Hence, the AIC values are examined to select the best fit distribution among the five. In this case, the lognormal distribution records the minimum AIC and therefore it is selected as the severity distribution.

5. Other: Asset Management (AM) + Corporate Finance (CF) + Retail Banking (RB) + Trading & Sales (TS)

Table 33 presents the total records for Other in 2009 as well the number of losses in body, tail, and low. In this case, fitting distribution in the tail is discarded since it contains only five observations.

Table 33: Total, Body, Tail, and Low Count of Losses of Other for 2009

Set	Total	Body	Tail	Low
2009_Other	57	29	5	23

Figure 16 presents the histogram of the frequency of gross loss amounts in Other for 2009.

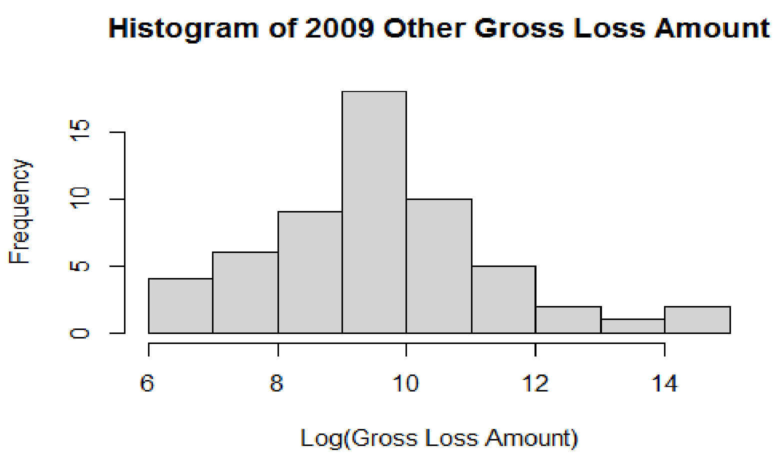


Figure 16: Histogram of the Frequency of Gross Loss Amounts in Other for 2009

Figure 17 presents the Q-Q and P-P plots respectively.

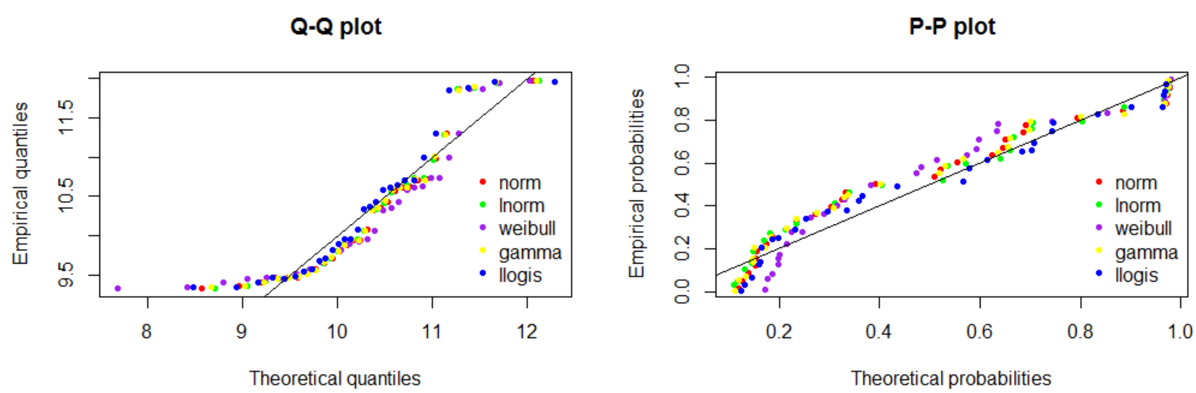


Figure 17: Q-Q and P-P Plots of the Fitted Severity Distributions of Other for 2009

An examination of the Q-Q and P-P plots shows that all five distributions are a good fit; that is based on their closeness to the best fit line.

Table 34 presents the estimated parameters of the fitted distributions in addition to the AIC and p-value of the KS test.

Table 34: Parameters of the Fitted Distributions, AIC, and p-value of Other for 2009

Distribution	Parameter 1	Parameter 2	AIC	p-value	Accepted
Normal	10.31	0.83	75.24	0.48	TRUE
Lognormal	2.33	0.08	73.63	0.54	TRUE
Weibull	12.17	10.71	81.33	0.32	TRUE
Gamma	161.07	15.62	74.13	0.52	TRUE
Log logistic	21.83	10.21	75.10	0.70	TRUE

The KS test results indicate that all distribution have a p-value > 0.05. Hence, the AIC values are examined to select the best fit distribution among the five. In this case, the lognormal distribution records the minimum AIC and therefore it is selected as the severity distribution.

Risk Category Severity Distribution Fitting

1. Clients, Products, & Business Practices (CPBP)

Table 35 presents the total records for CPBP in 2009 as well the number of losses in body, tail, and low. Fitting distribution in the tail is discarded since it contains only one observation.

Table 35: Total, Body, Tail, and Low Count of Losses of CPBP for 2009

Set	Total	Body	Tail	Low
2009_CPBP	100	39	1	60

Figure 18 presents the histogram of the frequency of gross loss amounts in CPBP for 2009.

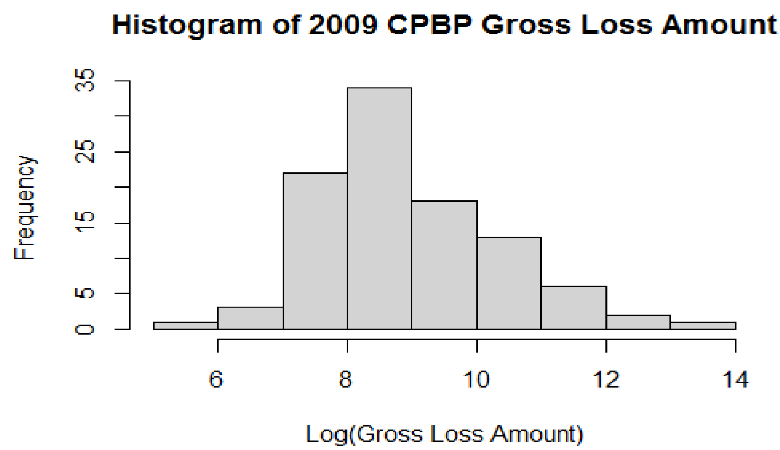


Figure 18: Histogram of the Frequency of Gross Loss Amounts in CPBP for 2009

Figure 19 presents the Q-Q and P-P plots respectively.

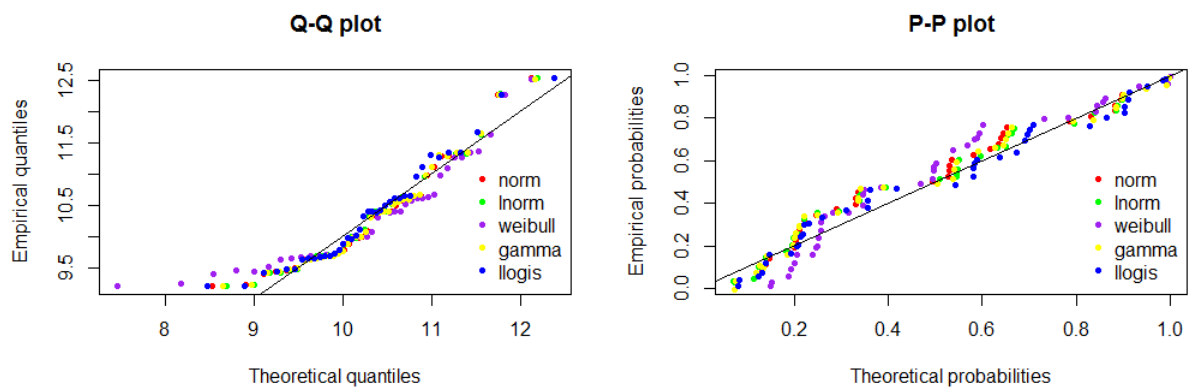


Figure 19: Q-Q and P-P Plots of the Fitted Severity Distributions of CPBP for 2009

An examination of the Q-Q and P-P plots shows that all five distributions are a good fit; that is based on their closeness to the best fit line.

Table 36 presents the estimated parameters of the fitted distributions in addition to the AIC and p-value of the KS test.

Table 36: Parameters of the Fitted Distributions, AIC, and p-value of CPBP for 2009

Distribution	Parameter 1	Parameter 2	AIC	p-value	Accepted
Normal	10.33	0.80	97.56	0.54	TRUE
Lognormal	2.33	0.08	95.29	0.61	TRUE
Weibull	12.00	10.72	107.57	0.19	TRUE
Gamma	171.70	16.62	95.99	0.59	TRUE
Log logistic	22.92	10.24	96.33	0.77	TRUE

The KS test results indicate that all distribution have a p-value > 0.05. Hence, the AIC values are examined to select the best fit distribution among the five. In this case, the lognormal distribution records the minimum AIC and therefore it is selected as the severity distribution.

2. Internal Fraud (IF)

Table 37 presents the total records for IF in 2009 as well the number of losses in body, tail, and low. In this case, fitting distribution in the tail is discarded since it contains only two observations.

Table 37: Total, Body, Tail, and Low Count of Losses of IF for 2009

Set	Total	Body	Tail	Low
2009_IF	43	22	2	19

Figure 20 presents the histogram of the frequency of gross loss amounts in IF for 2009.

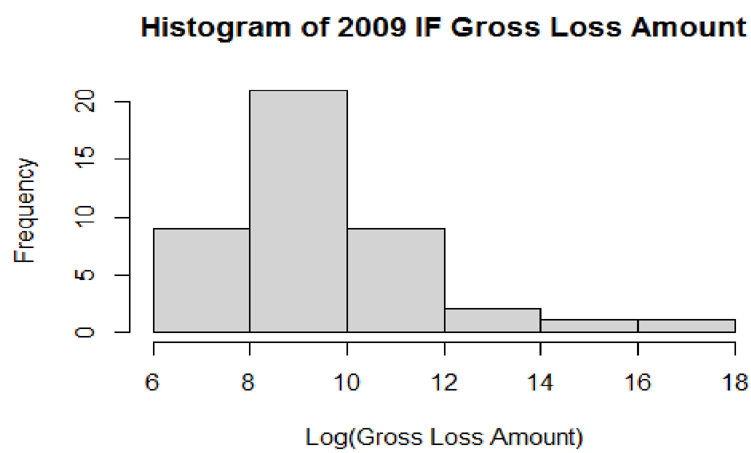


Figure 20: Histogram of the Frequency of Gross Loss Amounts in IF for 2009

Figure 21 presents the Q-Q and P-P plots respectively.

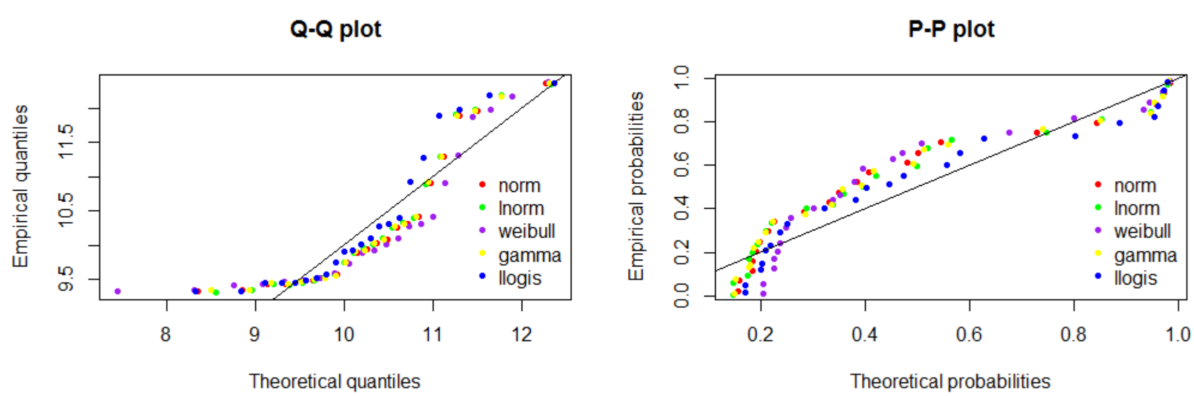


Figure 21: Q-Q and P-P Plots of the Fitted Severity Distributions of IF for 2009

An examination of the Q-Q and P-P plots shows that all five distributions are a good fit; that is based on their closeness to the best fit line.

Table 38 presents the estimated parameters of the fitted distributions in addition to the AIC and p-value of the KS test.

Table 38: Parameters of the Fitted Distributions, AIC, and p-value of IF for 2009

Distribution	Parameter 1	Parameter 2	AIC	p-value	Accepted
Normal	10.31	0.98	65.43	0.40	TRUE
Lognormal	2.33	0.09	63.60	0.50	TRUE
Weibull	10.18	10.79	70.31	0.20	TRUE
Gamma	117.11	11.35	64.18	0.47	TRUE
Log logistic	19.05	10.14	64.42	0.49	TRUE

The KS test results indicate that all distribution have a p-value > 0.05. Hence, the AIC values are examined to select the best fit distribution among the five. In this case, the lognormal distribution records the minimum AIC and therefore it is selected as the severity distribution.

3. Damage to Physical Assets (DPA)

Table 39 presents the total records for DPA in 2009 as well the number of losses in body, tail, and low. In this case, fitting distribution in the tail is discarded since it contains only two observations.

Table 39: Total, Body, Tail, and Low Count of Losses of DPA for 2009

Set	Total	Body	Tail	Low
2009_DPA	43	23	2	18

Figure 22 presents the histogram of the frequency of gross loss amounts in DPA for 2009.

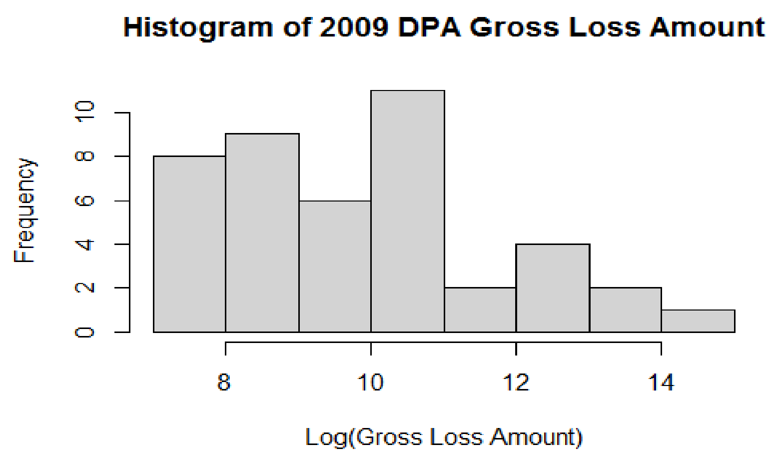


Figure 22: Histogram of the Frequency of Gross Loss Amounts in DPA for 2009

Figure 23 presents the Q-Q and P-P plots respectively.

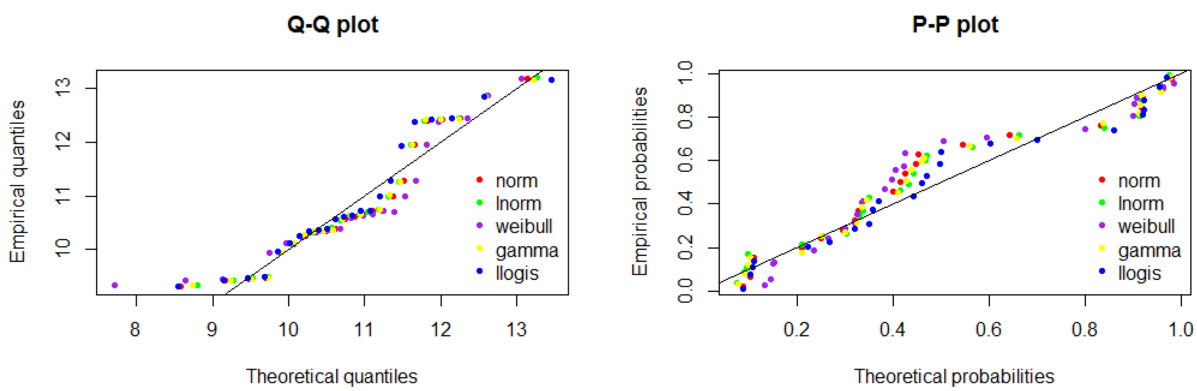


Figure 23: Q-Q and P-P Plots of the Fitted Severity Distributions of DPA for 2009

An examination of the Q-Q and P-P plots shows that all five distributions are a good fit; that is based on their closeness to the best fit line.

Table 40 presents the estimated parameters of the fitted distributions in addition to the AIC and p value of the KS test.

Table 40: Parameters of the Fitted Distributions, AIC, and p-value of DPA for 2009

Distribution	Parameter 1	Parameter 2	AIC	p-value	Accepted
Normal	10.86	1.12	74.69	0.27	TRUE
Lognormal	2.38	0.10	73.45	0.38	TRUE
Weibull	9.79	11.39	78.22	0.16	TRUE
Gamma	96.16	8.85	73.81	0.34	TRUE
Log logistic	16.82	10.72	74.74	0.61	TRUE

The KS test results indicate that all distribution have a p-value > 0.05. Hence, the AIC values are examined to select the best fit distribution among the five. In this case, the lognormal distribution records the minimum AIC and therefore it is selected as the severity distribution.

4. Rest: Execution, Delivery, & Process Management (EDPM) + External Fraud (EF) + Employment Practices & Workplace Safety (EPWS)

Table 41 presents the total records for Rest in 2009 as well the number of losses in body, tail, and low. In this case, fitting distribution in the tail is discarded since it contains only one observation.

Table 41: Total, Body, Tail, and Low Count of Losses of Rest for 2009

Set	Total	Body	Tail	Low
2009_Rest	51	28	1	22

Figure 24 presents the histogram of the frequency of gross loss amounts in Rest for 2009.

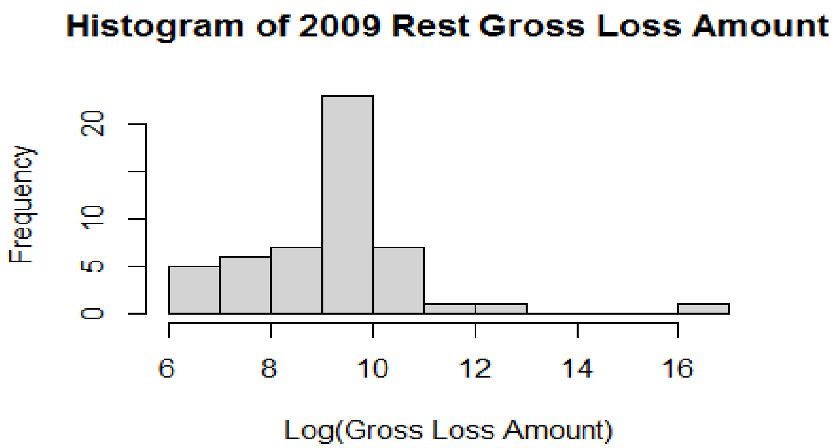


Figure 24: Histogram of the Frequency of Gross Loss Amounts in Rest for 2009

Figure 25 presents the Q-Q and P-P plots respectively

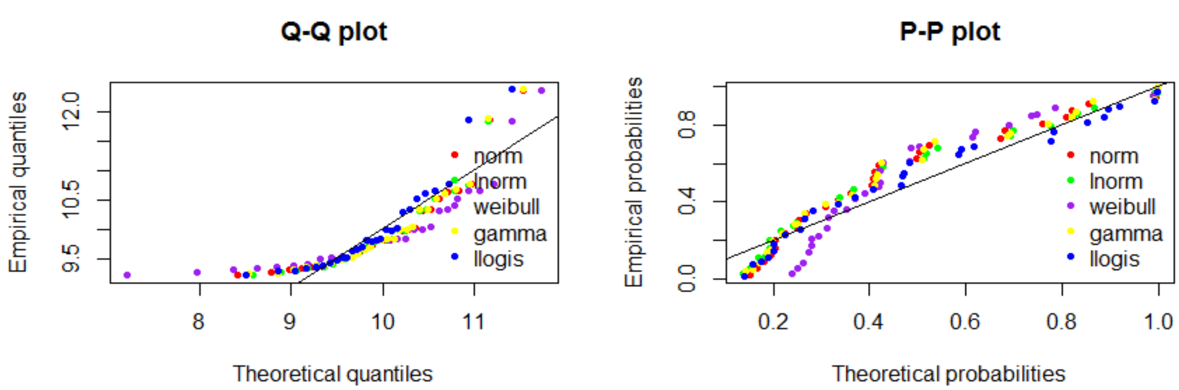


Figure 25: Q-Q and P-P Plots of the Fitted Severity Distributions of Rest for 2009

An examination of the Q-Q and P-P plots shows that all five distributions are a good fit; that is based on their closeness to the best fit line.

Table 42 presents the estimated parameters of the fitted distributions in addition to the AIC and p-value of the KS test.

Table 42: Parameters of the Fitted Distributions, AIC, and p-value of Rest for 2009

Distribution	Parameter 1	Parameter 2	AIC	p-value	Accepted
Normal	9.98	0.74	66.76	0.29	TRUE
Lognormal	2.30	0.07	63.61	0.30	TRUE
Weibull	11.09	10.35	78.52	0.07	TRUE
Gamma	194.41	19.49	64.62	0.28	TRUE
Log logistic	27.27	9.85	60.80	0.60	TRUE

The KS test results indicate that all distribution have a p-value > 0.05. Hence, the AIC values are examined to select the best fit distribution among the five. In this case, the log logistic distribution records the minimum AIC and therefore it is selected as the severity distribution.

Table 43 summarizes the severity distribution results for year 2009, whereby the selected best fit distribution along with its parameters are indicated for each business line and risk category.

Table 43: Summary of Business Line and Risk Category Severity Distributions for 2009

Set	Severity Distribution	Parameter 1	Parameter 2
AG	Lognormal	2.33	0.08
CB	Lognormal	2.33	0.10
RBR	Lognormal	2.31	0.05
PS	Lognormal	2.33	0.09
Other	Lognormal	2.33	0.08
CPBP	Lognormal	2.33	0.08
IF	Lognormal	2.33	0.09
DPA	Lognormal	2.38	0.10
Rest	Log logistic	27.27	9.85

The results show that all risk classes follow a lognormal distribution except for the "Rest" class which follows a log logistic distribution. The mean and standard deviation of all classes following a lognormal distribution are very close.

4.2.6 Convolution

Convolution is used to produce the aggregate loss distribution. That is, to compound the frequency and severity distributions. In this study, the Monte Carlo simulation is implemented, with number of simulations set to 10000, to estimate the aggregate loss distribution of each risk class. Regarding frequency, all distributions follow a Poisson distribution as addressed in section 4.2.4. The lambda parameter for each risk class is determined using the Poisson process as detailed earlier. Regarding severity, the distribution and respective parameters of each risk class are addressed in section 4.2.5. The aggregate distribution obtained follows a normal distribution. Table 44 presents the mean and standard deviation for each risk class. Similarly, this process is implemented for each year.

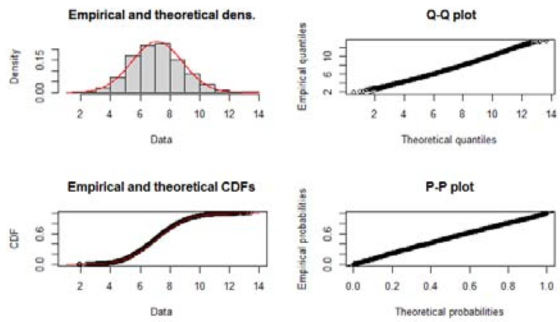
Table 44: Summary of Aggregate Distribution Parameters of All Risk Classes for 2009

Set	Frequency	Severity	Mean	SD
AG	Poisson	Lognormal	7.13	1.69
CB	Poisson	Lognormal	3.78	1.42
RBR	Poisson	Lognormal	4.70	1.56
PS	Poisson	Lognormal	7.17	2.16
Other	Poisson	Lognormal	15.57	2.37
CPBP	Poisson	Lognormal	30.42	2.87
IF	Poisson	Lognormal	5.39	1.59
DPA	Poisson	Lognormal	8.91	2.07
Rest	Poisson	Log logistic	0.08	0.02

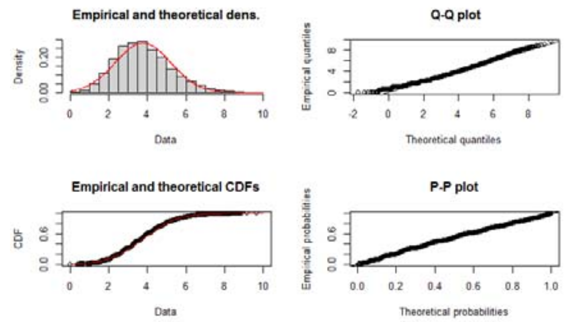
For each business line and risk category, the results of the convolution are presented in Figures 26 and 27. The bell shaped curve of the aggregate distributions are clearly observed in the density plots and this is also confirmed by the Q-Q, and P-P plots.

Business Line Aggregate Distribution Fitting

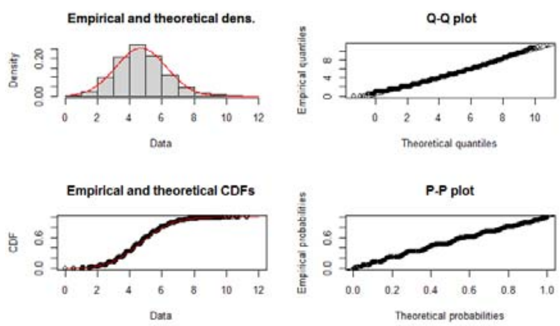
AG



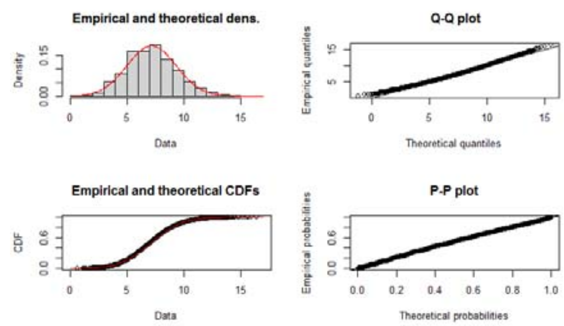
CB



RBR



PS



Other

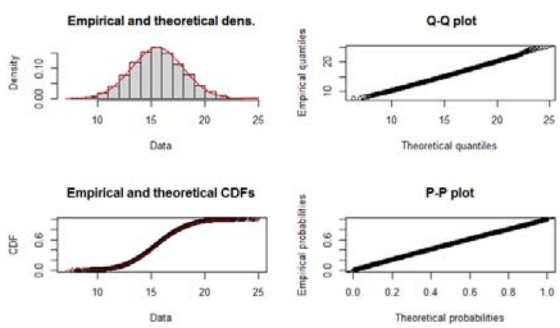
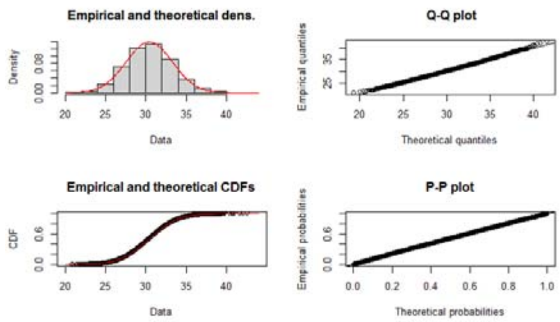


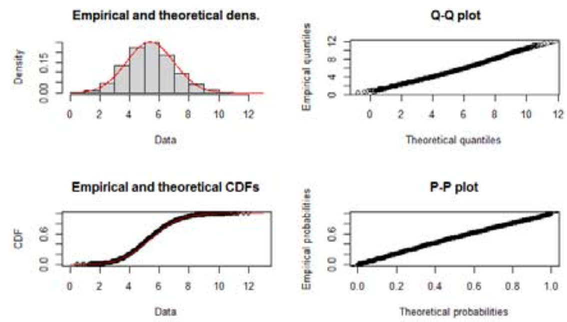
Figure 26: Density Function, CDF, Q-Q, and P-P Plots BL Aggregate Distributions for 2009

Risk Category Aggregate Distribution Fitting

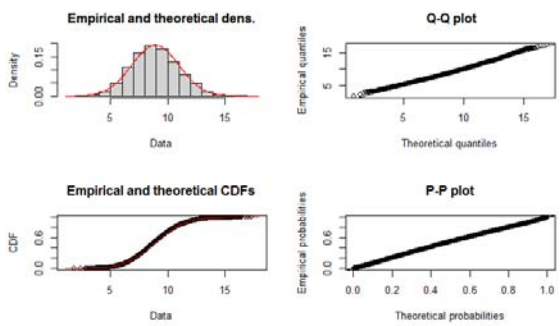
CPBP



IF



DPA



Rest

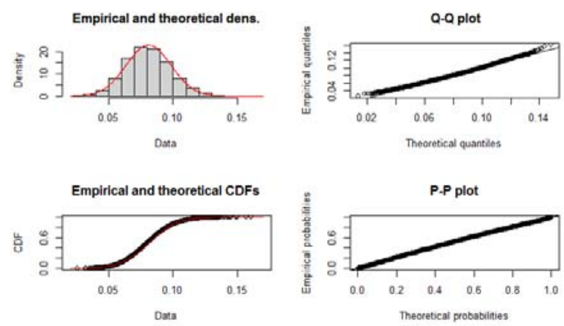


Figure 27: Density Function, CDF, Q-Q, and P-P Plots RC Aggregate Distributions for 2009

4.2.7 Copula

From the aggregate distributions of each risk class, the overall loss distributions can be obtained through a copula function. The t-copula, which is typically used in operational risk modeling, is implemented. Tables 45 and 46 present the results of the correlation matrix for the year 2009.

Table 45: Correlation Matrix Results of Business Lines for 2009

Business Line	AG	CB	RBR	PS	Other
AG	1	-0.102	0.072	0.486	-0.111
CB	-0.102	1	0.012	-0.049	-0.112
RBR	0.072	0.012	1	0.049	-0.079
PS	0.486	-0.049	0.049	1	-0.128
Other	-0.111	-0.112	-0.079	-0.128	1

Table 46: Correlation Matrix Results of Risk Categories for 2009

Risk Category	CPBP	IF	DPA	Rest
CPBP	1	-0.085	-0.074	-0.027
IF	-0.085	1	-0.015	-0.038
DPA	-0.074	-0.015	1	-0.059
Rest	-0.027	-0.038	-0.059	1

The number of simulations is set to 1,000,000 and the VaR is determined at the 99.9th percentile confidence interval as per the Basel requirements.

4.2.8 Capital Charge

Finally, the capital requirement is calculated for each year as the difference between the 99.9% VaR and the Expected Loss (EL). EL was calculated as the median (50th percentile), since the average is sensitive to extreme values. Capital requirement is calculated first using business line breakdown and then using risk category breakdown of the loss data.

Table 47 presents the VaR and capital requirement values for both business line and risk category methods for each year. For each year, significant differences are identified between the capital requirements obtained by the two methods.

Table 47: VaR and Capital by Business Line and Risk Category for All Years

Year	BL VaR	BL Capital	RC VaR	RC Capital
2009	58.8	642,688.1	51.5	110,977.2
2010	38.7	39,595.37	40.2	60,147.54
2011	109	3,850,904	244	12,168,160
2012	269	23,064,948	289	28,372,603
2013	36.6	38,057.07	34.8	31,485.27
2014	89.7	17,230,244	186	12,754,811
2015	24	8,314.57	51.7	21,966.54
2016	60.7	39,822.86	32.4	11,836.07
2017	65.9	1,003,365	45.2	44,166.93
2018	48.6	550,570.2	85.3	499,405.1

A proper understanding of the capital charge requires a detailed examination of the loss data. Figure 28 presents the variation of the gross loss amount over the 10 years. It can be clearly noted that the gross loss amount varies between 24$ and 55$ million between the years 2009 and 2015. In 2016, it decreases remarkably to around 9$ million, nevertheless it then increases back to 29$ million and records a tremendous surge to around 155$ million in 2018.

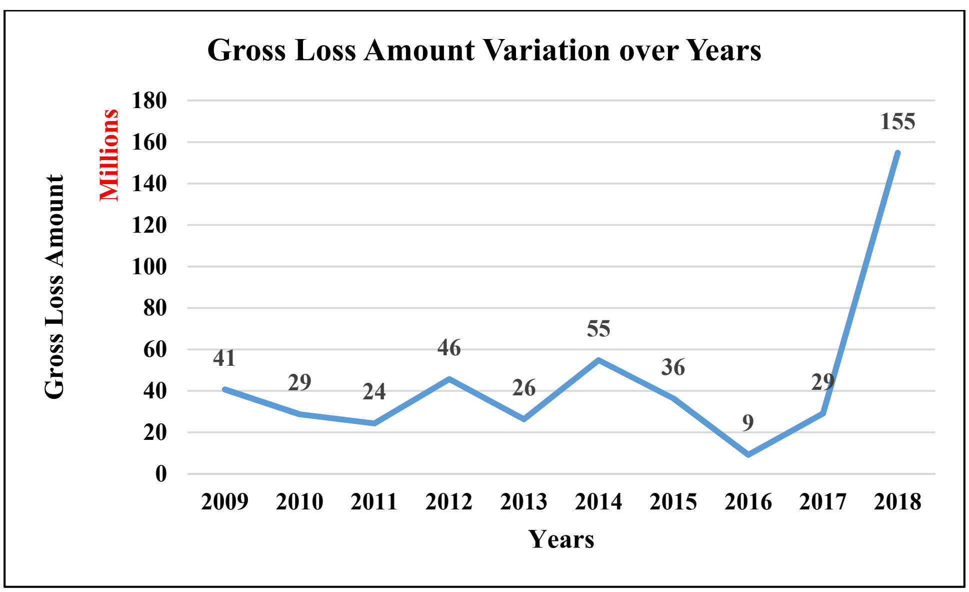


Figure 28: Gross Loss Amount Variation over Years

A detailed examination of the variation of gross loss amounts is required at two levels, by business line and by risk category as presented in Figures 29 and 30 respectively.

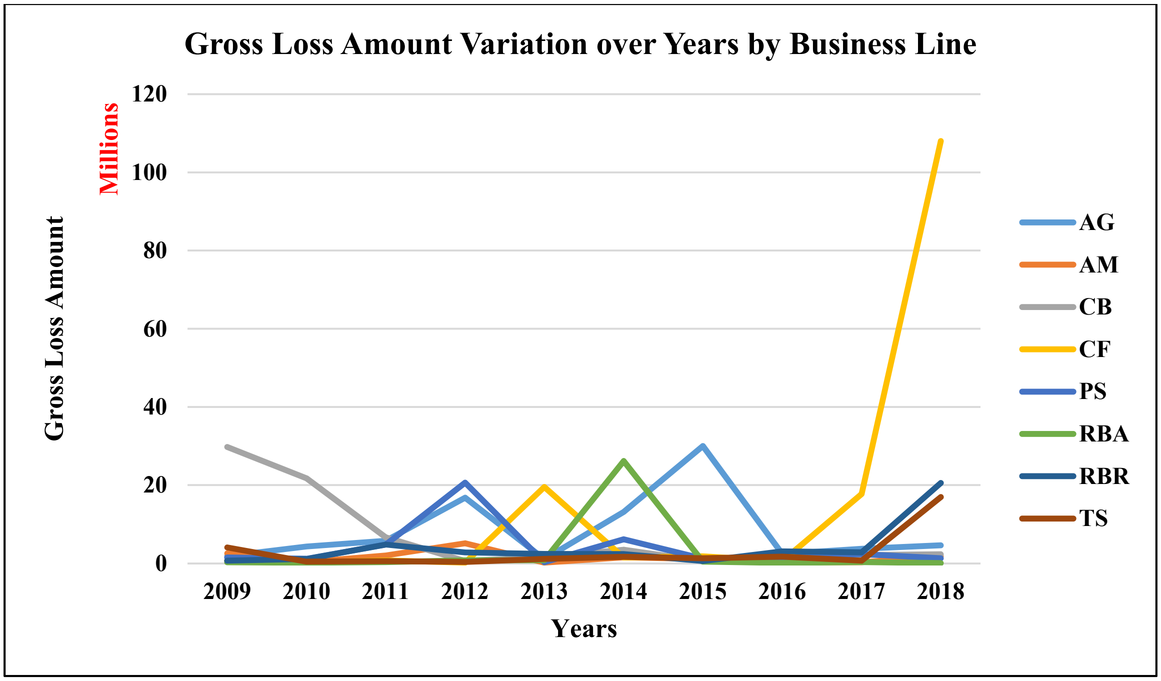


Figure 29: Gross Loss Amount Variation over Years by Business Line

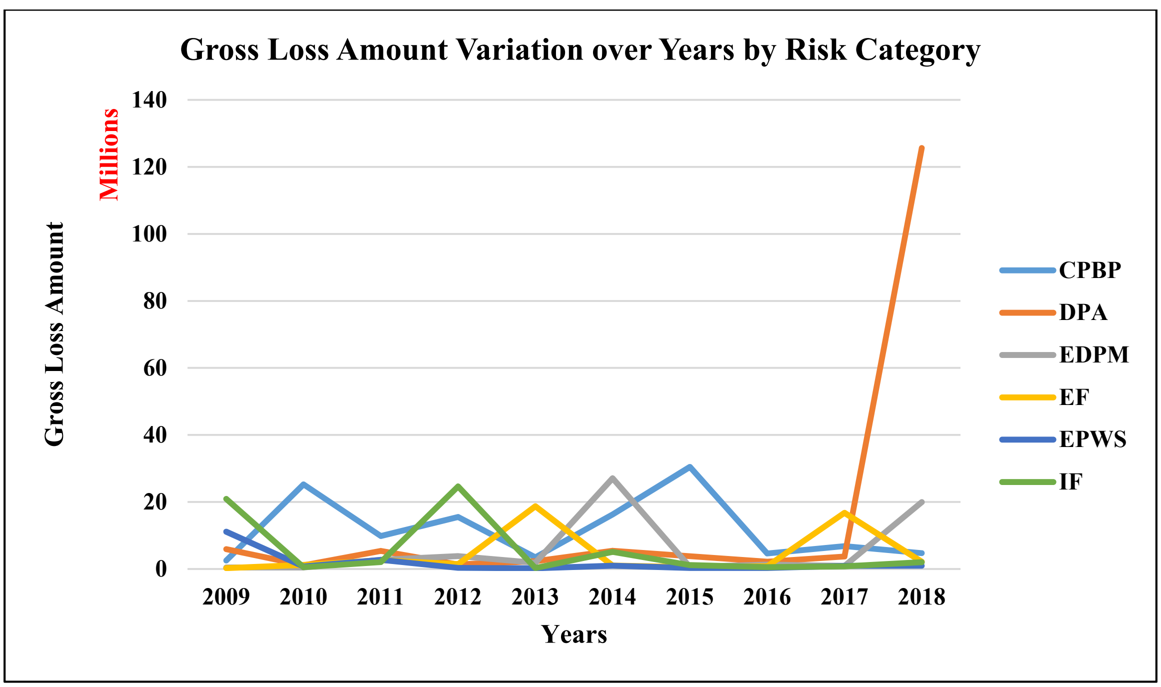


Figure 30: Gross Loss Amount Variation over Years by Risk Category

Gross loss amounts in business lines CF, PS, RBA, and RBR vary vastly over the years. In addition, an immense surge is noticed in CF from 2017 till 2018. Also, gross loss amounts in risk categories fluctuate greatly except for EPWS which is rather stable over the years. CPBP records the greatest fluctuations, however, similar to business line CF, an enormous surge is noticed in DPA from 2017 till 2018.

These variations have a huge impact on the capital requirement mainly because operational risk data involves high-frequency low-severity and low-frequency high-severity events. This is why operational risk loss data is characterized by a fat tail. These high severity events in the tail will dramatically impact the overall loss distribution and consequently the capital requirement.

As per Basel requirements, internally generated operational risk measures used for regulatory capital purposes must be based on a minimum five-year observation period of internal loss data. In this research study, the total 10 year period is considered, and a weighted average of the capital requirement is calculated. Table 48 presents the weighted average of the capital charge for both business line and risk category methods. Even though the capital charges varied significantly between these two methods per year, weighted averages over the 10 year period shows that both methods yield rather close capital charges. The business line method records a lower capital charge as compared to the risk category method. It is less by around 15%. Ultimately, and to diminish the impact of operational risk, the larger capital charge of 8,738,614$ is recommended for the next year.

Table 48: Weighted Average of Capital Charge for Business Line and Risk Category Methods

Over 10 Year Period	Business Line Method	Risk Category Method
Weighted Average of Capital Charge	7,407,347	8,738,614

Chapter Five: Conclusion and Recommendations

5.1 Conclusion

The Basel Committee defines operational risk as the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events. Operational risk is one of the main risk components for banks and financial institutions; as it is estimated to constitute between 15 and 25 % of total risks. Since the 1990s, the Basel Committee has been providing frameworks for the quantification of operational risk to allow banks cover such risks by required capital charges.

The aim of this research is to model operational risk under the Loss Distribution Approach. Due to the inaccessibility of real operational risk data, simulated data was used in this study. The data used included 3,192 operational risk events for a period of 10 years categorized by Basel's business line and risk category levels. However, the main obstacle in this study was that many risk classes lacked loss events and numerous other classes included a small number of events. Hence, many setbacks were imposed in frequency and severity distribution fitting. Furthermore, implementing the LDA was impossible at the level of business line and risk category combinations. As such, two main methods were considered. The LDA was implemented first using business line breakdown and second risk category breakdown of loss data. The capital charge was calculated for each method per year by following Basel standards and criteria.

The results show that capital charges vary significantly between the two methods adopted per year. However, in considering the 10 year period and calculating the weighted average, the two methods yield close values. The business line method records a lower capital charge by around 15% of that obtained by the risk category method. Consequently, the larger capital requirement of 8,738,614$ is recommended for the following year.

5.2 Recommendations for Future Work

Since any research study is defined by its scope of work, not all related questions can be answered. This research aimed at modeling operational risk under the Loss Distribution Approach. Two main procedures were implemented, the first using business line breakdown and the second using risk category of the breakdown. Certainly, the optimal method would be to consider combinations of business line and risk category cells. This was not possible in this research due to major restrictions in the data used. From this consideration, future research could include the following:

- Applying the R codes generated for this research on a different dataset, particularly data obtained from a database reflecting actual operational risk loss events.
- Adjusting the R codes generated for this research on another dataset covering the entire 56-cell matrix at the level of business line and risk category combinations.
- Modeling operational risk under the LDA with additional multivariate copulas such as the Gaussian or Gumbel copulas.
- Investigating supplementary methods under the AMA such as Bayesian approaches, neural networks, or Fuzzy modeling.
- Exploring Basel IV changes on capital charges which are due for future implementation.

References

2006. *International convergence of capital measurement and capital standards*. Basel: Bank for International Settlements.

2011. *Principles for the Sound Management of Operational Risk*. Basel: Bank for International Settlements.

2013. *Operational risk modelling framework.* Milliman Research Report.

2017. *Basel III: Finalising Post-Crisis Reforms.* Basel: Bank for International Settlements.

Angela, C., Bisignani, R., Masala, G., and Micocci, M. (2009). Advanced operational risk modelling in banks and insurance companies. *Investment Management and Financial Innovations,* 6(3).

Arindam Chaudhuri, & Ghosh, S. K. (2016). *Quantitative modeling of operational risk in finance and banking using possibility theory*. Springer International Publishing.

Bajaj, R. V. (2015). Operational risk capital estimation under BIA and TSA: a study of public sector and private sector banks in India. *DECISION*, *43*(1), 67–92.

Core.ac.uk. 2012. *Kernel Alternatives to approximate Operational Severity Distribution: An Empirical Application*. [online].

Couto, G., 2008. *Basel II: Operational Risk Measurement in the Portuguese Banking Sector and an Evaluation of the Quantitative Impacts*. [online]. Academia.edu.

Embrechts, P. (2009). Quantifying regulatory capital for operational risk.

Esterhuysen, J., Styger, P., & Van Vuuren, G. W. (2012). Calculating operational value-at-risk (OpVaR) in a retail bank. *South African Journal of Economic and Management Sciences*, *11*(1), 1–16.

Frachot, A., Georges, P., & Roncalli, T. (2001). Loss Distribution Approach for Operational Risk. *SSRN Electronic Journal*.

Fritz-Morgenthal, S. (2015). *A brief history of Operational Risk Management*. London.

Gregoriou, G. N. (2009). *Operational risk toward Basel III: best practices and issues in modeling, management and regulation*. John Wiley & Sons.

Han, J., Wang, W., & Wang, J. (2015). POT model for operational risk: Experience with the analysis of the data collected from Chinese commercial banks. *China Economic Review*, *36*, 325–340.

Laurent Balthazar. (2006). *From Basel 1 to Basel 3*. London Palgrave Macmillan UK.

Lin, T. T., Lee, C.-C., & Kuan, Y.-C. (2011). The optimal operational risk capital requirement by applying the advanced measurement approach. *Central European Journal of Operations Research*, *21*(1), 85–101.

Lu, J., Guo, L., & Liu, X. (2013). Measuring the operational risk of Chinese commercial banks using the semilinear credibility model. *The Journal of Operational Risk*, *8*(2), 3–34.

Momen, O., Kimiagari, A., & Noorbakhsh, E. (2012). Modeling the operational risk in Iranian commercial banks. *Journal of Industrial Engineering International*, *8*(1).

Moosa, I. A. (2008). A critique of the advanced measurement approach to regulatory capital against operational risk. *Journal of Banking Regulation*, *9*(3), 151–164.

Power, M. (2005). The invention of operational risk. *Review of International Political Economy*, *12*(4), 577–599.

Rahman, K. N., Black, D. A., & McDonald, G. C. (2014). An Application of Bayesian Inference on the Modeling and Estimation of Operational Risk Using Banking Loss Data. *Applied Mathematics*, *05*(06), 862–876.

Shevchenko, P., & Peters, G. (2013). Loss distribution approach for operational risk capital modelling under Basel II: Combining different data sources for risk estimation. *Journal of Governance and Regulation*, *2*(3).

Nuugulu, SM., & Kock, G. (2020) On a Loss Distribution Approach for Allocating Operational Risk Capital in Banks. *Journal of Mathematical and Statistical Analysis.* 3: 101.

Tharwat, A., A. ZeinEldin, R., Khalifa, H. A. W., & Saleim, A. M. (2018). Improving Losses Distribution Approach for Operational Risk using Fuzzy Numbers. *International Journal of Computer Application, 3*(8).

Vukovic, O. (2015). Operational Risk Modelling in Insurance and Banking. *Journal of Financial Risk Management*, *04*(03), 111–123.

Wang, Y., Li, J., & Zhu, X. (2017). A Method of Estimating Operational Risk: Loss Distribution Approach with Piecewise-defined Frequency Dependence. *Procedia Computer Science*, *122*, 261–268.

Wei, L., Li, J., & Zhu, X. (2018). Operational Loss Data Collection: A Literature Review. *Annals of Data Science*, *5*(3), 313–337.

Zhu, X., Li, J., & Wu, D. (2019). Should the Advanced Measurement Approach for Operational Risk be discarded? Evidence from the Chinese Banking Industry. *Review of Pacific Basin Financial Markets and Policies*, *22*(01), 1950007.

Zhu, X., Li, J., Chen, J., YangHuo, Y., Gao, L., Feng, J., Wu, D., & Xie, Y. (2014). A Nonparametric Operational Risk modeling approach based on Cornish-Fisher Expansion. *Discrete Dynamics in Nature and Society*, *2014*, 1–8.