UNIVERZITET U NOVOM SADU
PRIRODNO-MATEMATIČKI FAKULTET
DEPARTMAN ZA MATEMATIKU I INFORMATIKU

Ivana Manić

Mathematical Models for Estimation of Operational Risk and Risk Management

- master thesis -

Novi Sad, 2007.

”Pogledaj u sunce i tada će senke biti iza tebe”

Biography

I was born on March, 30th 1980 in Čačak, Serbia, and since 1991 I live in Novi Sad, Serbia. In 1999 I enrolled at undergraduate program in mathematics at the Department of Mathematics and Informatics, Faculty of Science, University of Novi Sad.

In 2004 I finished the studies with average mark 9,12/10 and became Bachelor of Science in Mathematics - Financial Mathematics.

The same year I started the master studies at Faculty of Science, University of Novi Sad and obtained a scholarship from Ministry of Science, Republic of Serbia. In the year of 2005 I was at the Department of Mathematics "F. Enriques", University of Milan, Italy for one semester. My stay at University of Milan was financed by European Commission through Tempus Project CD JEP 17017-2002 Mathematics Curricula for Technological Development.

During the master studies I participated at "ECMI Mathematical Modelling Week"; in 2005 at University of Novi Sad and in 2006 at Technical University of Copenhagen, Denmark. Further, I have attended the course "Financial Mathematics" in Plovdiv, Bulgaria supported by DAAD. In September 2006 I have presented my work in the field of operational risk at the XVI Conference on Applied Mathematics in Kragujevac, Serbia. In 2007 I have participated in project Modelling and Forcasting Stock Process Behaviour - Analysis of High Frequency Data at the Department of Mathematics and Informatics, University of Novi Sad. The project was conducting for Dresdener Kleinwort Securities Limited London. In July 2007 I have been at Credit Risk Training in VUB bank Bratislava, Slovakia organized and financed by Panonska banka ad Novi Sad.

During 2007 I worked in "M&V Investments" brokerage house in Belgrade, Serbia as a financial analyst. Since December 2007 I work in UniCredit Bank, Belgrade as risk controller.

Novi Sad, 22.11.2007. Ivana Manić

Preface

Motivation

In the field of banking supervision and risk management The Basel Committee on Banking Supervision (BCBS) established in 1974 takes the most important place. In 1988, the BCBS has issued the first accord, that is The Basel Capital Accord (Basel I) a framework for governing capital adequacy for internationally active banks. After a while, in 1999 the consultative version of The New Capital Accord (Basel II) was published, while the advanced version of Basel II was issued in 2006. The Basel Committee with its' Accords establishes standards for measuring the risk and the minimum capital requirements which should best suited to banks' actual risk exposures.

Expectations for full implementation of the Basel II was at the end of year 2006, while the advanced versions of the rules will be implemented by the end of 2007. In light of that, financial institutions around the world are increasing their focus on risk management to ensure they are well prepared.

With New Basel Accord II, operational risk was introduced as equally important as market and credit risks are, and defined as a risk of loss resulting from inadequate or failed internal processes, people, systems or from external events. Also, within Basel II banks have got a freedom to develop theirs own internal models for estimation of operational risk in order to provide more risk-sensitive capital models.

Accordingly, the basic motivation for this paper was the research of mathematical models that can be used for estimation of operational risk in a financial institution.

Acknowledgement

The idea for this master thesis came from the Mathematical - Industrial Days held at the Department of Mathematics and Informatics, Faculty of Science, the

University of Novi Sad in July 2006. The problem of estimation of operational risk in a bank institution was introduced by Dragica Mihajlović M.Sc in Economy, executive manager of Panonska banka ad Novi Sad to who I am sincerely grateful for all information, useful discussions and meetings about this topic.

Specially, I would like to thank my supervisor prof Dr Nataša Krejić for a huge support during this work and also during my studies. I am more than grateful for receiving right guidelines, suggestions and advices considering both mathematics and real world.

Also, I thank my professors from the Department of Mathematics for all knowledge and information they provided to me.

Further, I own a big thanks to my father Milorad, mother Blanka, sister Jelena, my friends and my colleagues for being there for me.

§ § §

This master thesis is conducted under the Project 114006 of Ministry of Science, Republic of Serbia, and project Tempus CD JEP 17017-2002 Mathematics Curricula for Technological Development. From 2004 to 2007 I have received the scholarship from Ministry of Science, Republic of Serbia.

Outlines

In Chapter 1 the notation adopted in this paper is given, together with basic well known facts from econometrics and statistics.

Chapter 2 starts with the introduction of the Basel Committee and the Basel II Accord. Later, the operational risk is defined, and the complete operational risk management framework is explained.

Chapter 3 deals with capital models for operational risk allocation. These models calculate the capital charge for operational risk exposure and the basics of all three models are presented. More details are given for the Loss Distribution Approach.

Chapter 4 can be considered as the main chapter of this paper. It starts with descriptive analysis of loss data moving to parametrical fitting of frequency and severity of data. The attention was put on the parameters estimation of chosen distribution and consequently, the standard (Maximum Likelihood Estimation) and adjusted approach are explained. The Expectation-Maximization algorithm is adopted and explained in more details. Further, the statistical goodness-of-fit tests are presented for complete data set and left-truncated data sets. Since the operational risk loss can be extremely high value with the possibility to happen very small, Extreme Value Theory was considered. Two basic approaches are presented, i.e. Peaks over Threshold and Block Maxima. The final step in Loss Distribution Approach is the aggregation of both chosen distributions for severity and frequency of loss data. Monte Carlo simulation method is described and implemented for this model. The last part of Chapter 4 is the correlation effect among loss data. However, formulas for uncorrelated and correlated loss data are given in order to determine the capital charge for operational risk.

In Chapter 5 the empirical results obtained for a given data are reported. The organization of this chapter is the same as for Chapter 4.

At the end of the thesis the list of considered literature is listed.

Contents

List of Figures

Chapter 1

Introduction

1.1 Notation

In this paper the following notation is adopted.

The random variables are denoted with capital letters e.g. X, Y and its corresponding realizations (i.e. observations, random draw) with small letters x, y. If there is n-times realization of X then a small boldface letter presents a corresponding random vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$.

The cumulative distribution function (cdf) of X is denoted as $F_X(x;\theta)$ where θ is parameter or set of parameters. The corresponding probability function is $f_X(x;\theta)$. Sometimes the index is omitted when it is obvious from which distribution random variable comes, e.g. $F(x;\theta)$.

Convolution. Suppose that two independent random variables X and Y with their cdf F_X, F_Y and corresponding probability density functions (pdf) f_X, f_Y are given. If we observe the sum $U = X + Y$, then, the pdf f_U is the convolution integral

$$f_U(u) = \int_{-\infty}^{\infty} f_X(u-y) \cdot f_Y(y)dy = \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(u-x)dx \; .$$

The notation is $F_U = F_X \star F_Y$. The n-fold convolution is defined as $F^{1\star} = F$, $F^{n\star} = F^{(n-1)\star} \star F$.

Conditional expectation. Suppose that (X, Y) is a joint random variable explained by pdf $f_{(X,Y)}(x, y)$. Then, the conditional expectation for every

$g(\cdot, \cdot)$ function given $X = x$ can be expressed as

$$\mathrm{E}(g(X,Y) \mid X = x) = \int_{y \in \mathbf{Y}} g(x,y) f_{Y|X}(y \mid x) \, \mathrm{d}y \; . \tag{1.1}$$

Skewness and kurtosis. The skewness is the third standardized moment that explains the asymmetry of data around the mean value, while the forth standardized moment is the measure of tail-heaviness called kurtosis. Let X be the random variable with the mean μ and the standard deviation σ, then the formulas are

$$Skewness = \frac{E((X-\mu)^3)}{\sigma^3} \; ,$$

$$Kurtosis = \frac{E((X-\mu)^4)}{\sigma^4} \; .$$

For the normal distribution the skewness is equal to zero which means that positively skewed distributions have more probability mass concentrated on the left side of the mean, while, negatively skewed distributions have more probability mass concentrated on the right side. The kurtosis of normal distribution is equal to 3 and distributions with more heavier tails than the normal distribution, have kurtosis greater than 3. Respectively, distributions with ticker tails have kurtosis less than 3.

Quantile. The quantile (percentile) presents a cutoff value x such that the area to their right is a given probability $1 - \alpha$ where $\alpha \in [0, 1]$ expressed as

$$\begin{aligned} 1 - \alpha &= \mathrm{P}(X > x) = 1 - F(x) \\ &= \int_x^{+\infty} f(x) \mathrm{d}x \; . \end{aligned}$$

Therefore, the α^{th} quantile ($\alpha \times 100\%$) can be also defined as

$$x = \inf\{x : F(x) \geq \alpha\} \; .$$

Chapter 2

Operational Risk

2.1 The Basel Committee, The Basel Accords

The world's most important organization in the field of banking supervision and risk management is The Basel Committee on Banking Supervision (BCBS). It was established by the central-bank governors of the Group of Ten countries[1] in 1974 as a part of the world's oldest international financial organization Bank of International Settlement (BIS). The aim of BCBS is to "formulate broad supervisory standards and guidelines and recommends statements of best practice in the expectation that individual authorities will take steps to implement them through detailed arrangements - statutory or otherwise - which are best suited to their own national systems", [4].

In other words, BCBS tends to find the best common approaches and common standards for every member country in order to promote the advancement of risk management in the banking system, strengthen banking supervisory frameworks and to improve financial reporting standards. To achieve this, BCBS has published many documents in the field of capital adequacy, banking problems, accounting and auditing, core principles for effective banking supervision, credit risk and securitization, market risk, operational risk, money laundering and terrorist financing, transparency and disclosure. For the risk management the most important documents are the Basel Accords, Basel I and Basel II.

In 1988, the BCBS has issued the first accord, that is The Basel Capital Accord (Basel I), a framework for governing capital adequacy for internationally active banks. After a while, in 1999 the consultative version of The New Capital

[1]The members of Group of Ten are Belgium, Canada, France, Germany, Italy, Japan, the Netherlands, Sweden, Switzerland, the United Kingdom, the United States of America.

Accord (Basel II) was published, and in 2004 its finale version. The advanced version of Basel II was issued in 2006, [3].

The Basel Committee with its' Accords establishes standards for measuring the risk and the *minimum capital requirements* which should best suited to banks' actual risk exposures.

Regulatory capital, also called the capital charge or the minimum capital requirement, is the capital defined by the regulators that bank should set aside as a buffer against its potential losses. The regulatory capital is meant to assure bank's ability to cover major potential losses (or to cover significant but not catastrophic losses) without causing a banking crisis. Consequently, regulatory capital management should ensure the soundness and the stability of the banking sector and protect depositors.

Economic capital is, on the other hand, every kind of capital (such as book capital, reserves, charges etc.) that can absorb economic losses without interrupting any banking activity. It is calculated according to the bank's experts opinions and it is not a subject of supervisory review. Further, the economic capital management helps in identifying the measure of risks, base strategic decisions on accurate information, strengthen an institution's long-term profitability and competitiveness. Indeed, the regulatory and economic capital are highly connected.

The universe of the risks which banks can face is composed of three basic types - credit, market and operational risk. In Basel I the main focus is on market and credit risk, leaving operational risk with no operational capital requirements. Basel II, as a big improvement of Basel I, has introduced the operational risk equally important as market and credit risks are. The risks are defined as following.

Credit risk is the risk that a counterpart will not be able to meet their contractual obligations for full value.

Market risk is the risk of losses in on- and off- balance sheet positions arising from movements in the level or volatility of market prices.

Operational risk is the risk of loss resulting from inadequate or failed internal processes, people, systems or from external events.

The aim of Basel II is to promote safety and soundness in the financial system. It means that it should make capital requirement sensitive to bank's risk, maintain current level of firm's capital, focus on international banks and apply comprehensive risk approach. These objectives can only be achieved through the mutually reinforcing of the Basel's three pillars, [3], as shown in Figure 2.1.

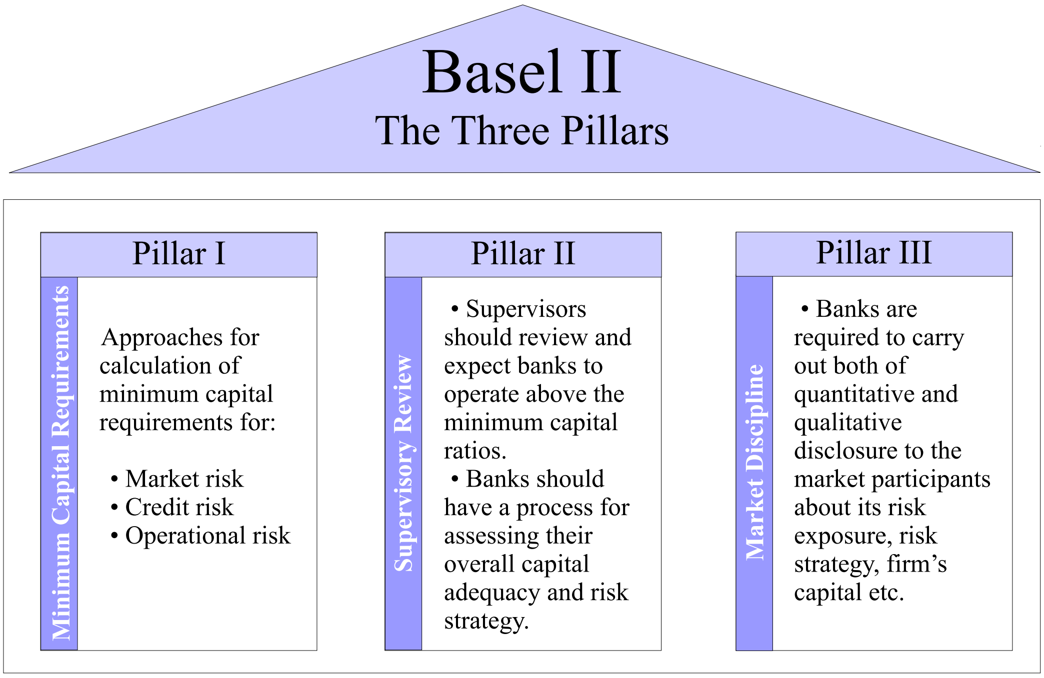

Figure 2.1: The Structure of Basel II Document

Pillar 1 - Minimum Capital Requirements. Each bank institution is required to calculate the minimum capital requirements, that is, a monetary caution that will be set aside as a reserve against the risks that bank takes. Pillar 1 is based on capital requirements defined in Basel I for credit and market risk, and also, on a new capital charge for operational risk. Furthermore, Pillar 1 sets several approaches to credit, market and operational risk estimation, increases risk sensitivity through more refined credit risk weights and internal ratings based approaches, and defines constituents of firm's capital.

Pillar 2 - Supervisory Review Process. According to Pillar 2 supervisors have more active role with possibility to take actions where necessary. They should review banks' internal capital assessments, ensure that banks operate above the minimum regulatory capital ratios and require rapid remedial actions if capital is not maintained or restored. The capital ratio is considered as bank's capital that should be available for potential losses caused by credit, market and operational risk. Currently, it is measured by

$$CapitalRatio = \frac{TotalCapital}{CreditRisk + 1.25(MarketRisk + OperationalRisk)}$$

Further, since the total capital ratio must not be lower than 8% the **total regulatory capital** is greater or equal to the following sum:

$$TotalCapital \geq 0.08 \cdot CreditRisk + MarketRisk + OperationalRisk$$

Supervisors have the obligations to make publicly available the criteria which is used in the process of risk management; in a transparent and accountable manner. Consequently, this review process encourages financial institutions to develop better risk management techniques in monitoring, measuring and managing risks.

Pillar 3 - Market Discipline. Financial institutions must carry out quantitative and qualitative disclosure to the market participants regarding firm's capital, risk management practices, structure and organization, risk assessment processes, risk exposures, and the capital adequacy of the institution. In general, this kind of disclosure reveals the level of risk management sophistication to the market. Additionally, under the Pillar 3 the volume and the frequency of reporting are increased.

Expectations for full implementation of the Basel II basic framework for every member country of BCBS was at the end of year 2006, while the advanced versions of the rules will be implemented by the end of 2007. In light of that, financial institutions around the world are increasing their focus on risk management (specially on operational risk management as relatively new and not enough explored form of risk) to ensure they are well prepared.

2.2 Defining Operational Risk

Until recently, there was no agreement on definition of operational risk (OR). The earlier definitions were either inconsistent or overlapping, going from "everything except market and credit risk" to "losses due to failures in the operational process". Finally, in 2001 the consensus on OR definition has been reached and BCBS [1] has defined OR as follows.

Definition 2.1. Operational risk is the risk of loss resulting from inadequate or failed internal processes, people, systems or from external events.

This definition includes legal risk[2], but excludes strategic and reputational risk.[3]

Causes	Business Line	Event Type	Effects
1. People	1. Corporate Finance	1. Internal Fraud	1. Loss of recourse
2. System	2. Trading & Sales	2. External Fraud	2. Write-down
3. Processing	3. Retail Banking	3. Employment Practices & Workplace Safety	3. Loss of physical asset
4. External causes	4. Commercial Banking	4. Clients, Products & Business Practices	4. Restitution
	5. Payment &Settlement	5. Damage to Physical Assets	5. Legal cost / settlement
	6. Agency Services	6. Business Disruption & System Failures	6. Loss of money
	7. Asset Management	7. Execution, Delivery & Process Management	7. Loss of important information etc.
	8. Retail Brokerage		
Example: **Cause** External causes	**BL** Retail Banking	**ET** External Fraud	**Effect** Loss of money

Table 2.1: Causes, Business Lines, Event Types and Effects

From the above definition it should be noted that OR highly depends on characteristics of each bank institution; its specific processes, personal, culture

[2]Legal risk is a risk of loss resulting from legal actions (i.e. fines, penalties, or punitive damages) and private settlements.

[3]The strategic risk is the risk of unexpected losses resulting from incorrect decisions taken by senior management while the reputational risk is the loss arising from a damaged firm's reputation.

and technology. Apparently, OR losses can appear from different types of events such as a management failure, inadequate procedures and controls, malfunction of IT system, poorly trained, overworked or unmotivated employees, unauthorized activity by employees, breaking of systems security by hacking damage or theft of information, external catastrophic events, earthquake, fire, or terroristic attack like one on the World Trade Center on September 11, 2001 etc. In fact, the main cause of the OR loss heavily depends on internal processes in the firm, business strategy, technology. Thus, this form of risk continuously changes with firm's development and competition.[4]

The analysis of OR losses can be performed according to causes, loss event types, effects that drive the OR losses, and, also by business lines where the OR loss occurred. See Table 2.1. In order to provide the general classification of OR losses BCBS [3] has defined four basic categories of the causes of OR, seven event types and eight business lines of bank's organizations[5].

The causes: people, processing, system and external factors.

Business Lines (BL): Corporate Finance, Trading & Sales, Retail Banking, Commercial Banking, Payment & Settlement, Agency Services, Asset Management and Retail Brokerage.

Event Types (ET): Internal Fraud, External Fraud, Employment Practices & Workplace Safety, Clients, Products & Business Practices, Damage to Physical Assets, Business Disruption & System Failures and Execution, Delivery & Process Management.

Effects: loss of recourse, write-down, loss of physical asset, restitution, legal cost and settlement, loss of money, loss of important information and etc.

As we will see in the following chapters, this classification of OR's loss types and business lines is highly useful in identifying, assessing, and also, in calculating the OR capital charge. Moreover, it is required for the banks under the Advanced Measurement Approaches.

It is worth noting that operational losses vary quite a lot between different business lines and event types. As an example of this, Table 2.2 and Table 2.3 present the percentages of the operational loss data disaggregated by BL and

[4]Regarding the nature of OR losses, BCBS has left, for each bank institution, the possibility to change the definition of OR according to their own individual characteristics.

[5]For more details see [3] Annex 8: Mapping of BL and Annex 9: Detailed Loss Event Type Classification

ET, respectively. All losses exceeded $1 million, occurred in the United States of America and have been collected by vendor OpVantage[6], [5].

BL		% of all losses
BL1	Corporate Finance	4 %
BL2	Trading & Sales	9 %
BL3	Retail Banking	39 %
BL4	Commercial Banking	16 %
BL5	Payment & Settlement	1 %
BL6	Agency Services	3 %
BL7	Asset Management	6 %
BL8	Retail Brokerage	22 %

Table 2.2: OR Losses by BL

ET		% of all losses
ET1	Internal Fraud	27 %
ET2	External Fraud	16.6 %
ET3	Employment Practices & Workplace Safety	3.3 %
ET4	Clients, Products & Business Practices	48.1 %
ET5	Damage to Physical Assets	0.3 %
ET6	Business Disruption & System Failures	0.4 %
ET7	Execution, Delivery & Process Management	4.2 %

Table 2.3: OR Losses by ET

One can immediately see that the business line with the most observations is Retail Banking (BL3) (39% of all losses) while the most frequent event type is Clients, Products and Business Practices (ET4) (48.1% of all losses). Yet, it is not necessary the case that the most frequent business line (or event type) has in the same time losses with the highest impact on the amount of lost money. It is more often that OR losses which are highly frequent are small/medium-sized losses, while the low-frequent ones are usually large ones.

2.3 Operational Risk Management Framework

Operational risk management framework is the process that involves the whole bank organization at all levels of management. The Basel Committee has defined the OR management framework in two documents, that is, Basel II [3] and Sound Practices for the Management and Supervision of Operational Risk [2]. There are quantitative and qualitative approaches to management of OR.

The Basel II has proposed three basic models for calculating operational capital charge: the Basic Indicator Approach (BIA), the Standardized Approach

[6]Namely, after the introduction of OR in 1999 with the first consultative version of Basel II banks have started to collect more information on their historical OR losses (creating internal loss databases). During the time, the data vendors like OpVantage, OpRisk Analytics and British Bankers Association have been formed in order to make easier exchange of operational loss experience among bank's institutions. What actually vendors do is collecting the available data from public information sources and making external database of OR losses.

(SA), and the Advanced Measurement Approaches (AMA). These capital models estimate the OR capital charge, that is, the OR regulatory capital and they differ among each other by the level of risk-sensitivity. The AMA model is the most risk-sensitive one and, until now, the most appropriate one for banks which tend to understand their OR exposure in order to find better control, monitor and mitigation of it. All capital models will be discussed in more details in the following chapters.

On the other hand, the Basel's document [2] has defined ten qualitative principles which banks should consider and adopt for their OR management framework. However, these principles are required for the banks under AMA capital model.

The principles are:

1. The board of directors should be **aware** of the major aspects of OR, **approve and periodically review** the OR management framework.

2. The board of directors should ensure that the framework is **subject to effective internal audit**.

3. Senior management has **responsibility** for implementing the framework, and all levels of staff should understand their responsibilities.

4. Banks should **identify** the OR in all products, activities, processes and systems for both existing operations and new products.

5. Banks should establish the processes to **regularly monitor** OR profiles and material exposure to losses.

6. Banks should have policies, processes and procedures to **control or mitigate** OR. They should assess the feasibility of alternative strategies and adjust their exposure appropriately.

7. Banks should have in place **contingency and business continuity plans** to ensure their ability to operate as going concerns in the event of business disruption.

8. Bank supervisors should require banks to have an **effective** OR management **strategy** as part of an overall approach to risk management.

9. Supervisors should conduct **regular independent evaluations** of the related bank OR management strategies.

10. Banks should make **sufficient public disclosure** to allow market participants to assess their approach to OR management.

According to the above principles the OR management framework has four basic components, that is, **strategy**, **process**, **infrastructure** and **environment**. See Figure 2.2.

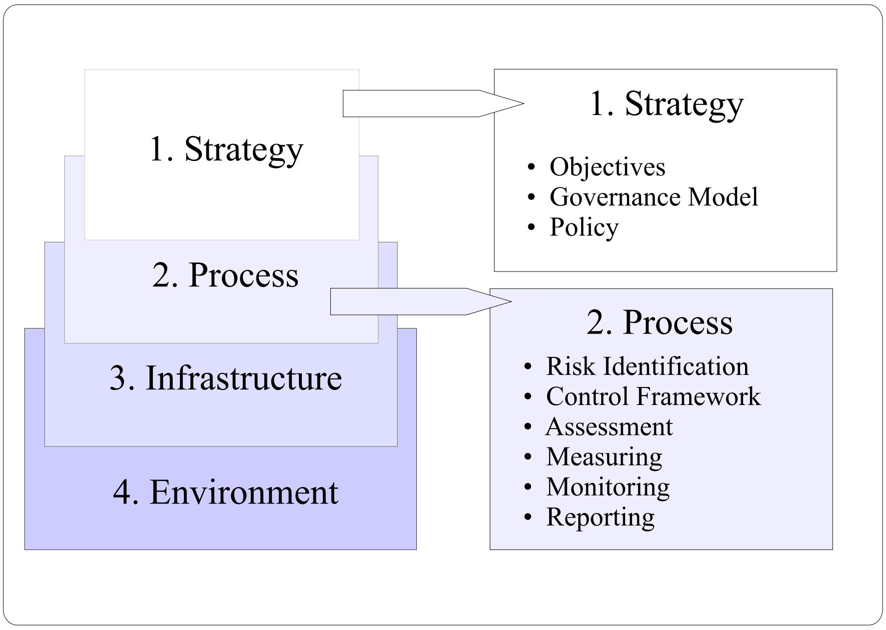

Figure 2.2: The OR Management Framework

- The whole OR management framework starts from the definition of bank's strategy. The strategy is composed of objectives, governance model and policy, and hence, it gives the basic form for the entire management framework. Namely, the strategy should define the goals and aims of the organization, the organization's risk appetite, the organization approach to the risk management, the policy of management and the responsibilities. The senior management forms the strategy while the board of directors approve it.

- The second phase of the OR management framework is the process. It contains six different parts; risk identification, control framework, risk assessment, measuring, monitoring and risk reporting. While the strategy is formed for the bigger time horizon, the process is more based on the daily activities and decisions which are performed within the strategy. Each part of the process will be explained later.

- Infrastructure should provide appropriate IT system, the database of OR losses and other tools which are used in the management process.
- The environment should define responsibilities for every part of the process, provide the appropriate bank's organization structure, set the communication, attitudes, practices in the bank on the right level, and everything else that characterize bank's daily risk activities. Additionally, environment should always monitor and assess external factors like industry trends, new regulators requirements, law, competitors' experiences etc. All this should be done in order to adjust, as necessary, the internal processes to the new situations.

Now, before explaining the stages of OR management process the following fact should be noted. In most cases, for banks that integrate for the fist time the OR management, it is very difficult to build the model which will perfectly fit to their organization structure and risk exposure. Namely, the problem is that all parts of the OR management framework can not be built with the same degree of quality at the same time. There are lots of limitations such as no history data available for measuring the risk, no trained staff for managing the process or there are no appropriate infrastructure and IT systems to be used. But, since banks need to have some kind of model for OR management, the common practice shows that, in the first stage of implementation of risk management, it is better to have model with reasonable characteristics than not having it at all. Of course, after building the first version of the OR management framework banks need to improve and adjust it all the time.

2.3.1 The OR Management Process

Perhaps the most important part of the management framework is the OR process which goes through six different parts beginning with risk identification, and followed by control framework, risk assessment, risk measurement, risk monitoring and finally risk reporting.

Risk Identification. The main question of the OR management framework is the identification of OR, that is, which types of risk exist and what is their influence on the banks' activities. Obviously, without appropriate identification it is very difficult to reach the successful management. The

Basel's definition of operational risk gives the global view of it, but only after collecting the internal (historical) data about OR losses one can be more familiar with potential causes and types of this risk. However, the Basel's definition can be modified and expanded. See footnote 4.

The tool that is used is the **risk mapping**. This process is carried out for every organizational unit in order to get the information on the unit's risk exposure, risk type and its corresponding degree. Basically, the risk's degree can be explained in terms of frequency and severity. The frequency is the number of loss events for some period of time, while the severity of risk is the impact of the event i.e. the lost amount of money. In risk mapping the severity and frequency can be expressed qualitatively (high, medium, low) or quantitatively.

The result of quantitative risk mapping is the probability-impact diagram i.e. the typical plot of expected loss frequency against severity for each risk event type or business line. Usually, the risk map is plotted on logarithmic scale. One typical risk map is shown in Figure 2.3.

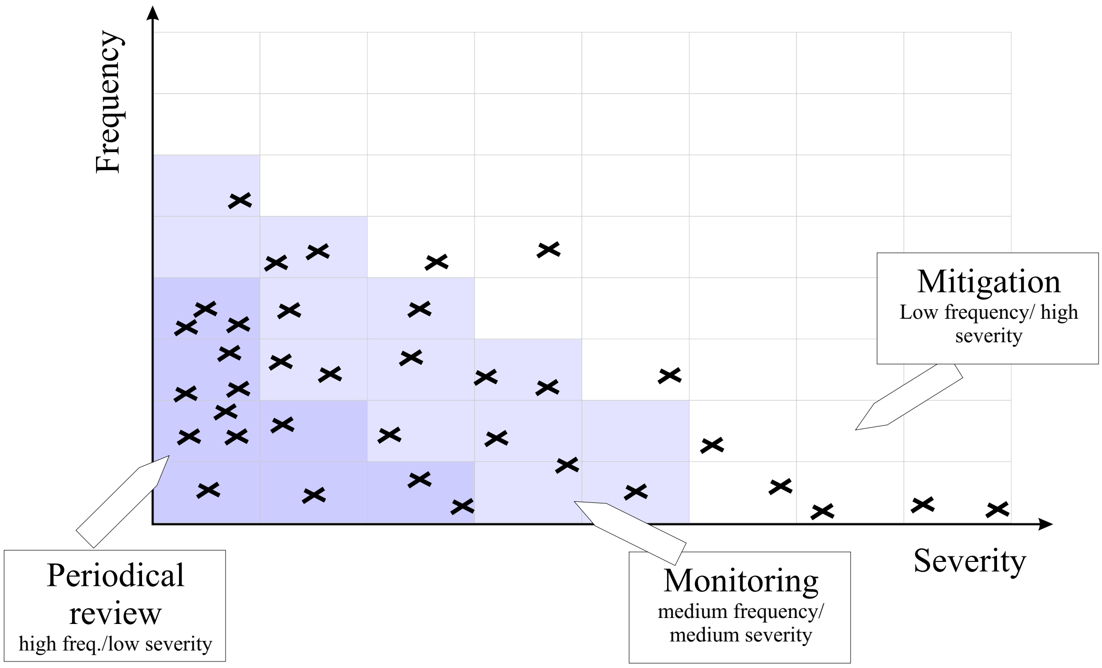

Figure 2.3: The Example of Risk Map

In the risk identification process one should also consider the external factor and industrial trends since some new form of OR can appear. Moreover, the old types of risk can change its severity and frequency.

Control Framework. The control framework defines the most appropriate approach to control every identified risk. In this stage it is consider how well some control approaches operating as well as the cost of its implementation. It is important to know which risk can be handled without extra costing and up to what degree. Some of control approaches are the following: periodically reviewed process, monitoring, mitigation and insurance.

From the risk map it can be seen which risk events should be in the main focus for the management control and under which type of control. In Figure 2.3 the dark shaded region presents high-frequency low-severity events which are usually covered by the cost of the business. Since they are not of a big influence on the OR measure (i.e. OR capital charge) they are only periodically reviewed. The shaded region is for the medium-frequency medium-severity loss events. These events form the basis for calculating the expected losses and they can be reduced with appropriate management control. They are under monitoring process. The low-frequency high-severity loss events are in the white region. They are considered as the unexpected losses and they are the most important ones. Their reduction is sometimes very difficult to obtain and, hence, they are subject of mitigation. Of course, the detailed analysis, insurance polices and planning are necessary for their appropriate OR control. The very low-frequency and very high-severity events are catastrophic losses and risk capital charge is not constructed to cover them. Namely, these losses are under insurance policy since firm can not deal with these risk exposures by itself.

Additionally, the control framework must be in line with the defined strategy. Its characteristics should not prevent the achievement of basic firm's objectives.

Risk Assessment. Risk assessment is the qualitative process that results in obtaining the risk profile of the whole financial organization and the appropriate action plan. The action plan describes responsibilities and the actions that bank should perform. Hence, it is the subject of approval of the senior management. To be more precise, risk assessment needs to determine and assess the risk exposure of the firm, how well the control and monitoring of the risk is carried out, what are the weaknesses of the management framework, which actions should be done in order to improve the risk management and also define the responsibilities.

The most common tool is the **self-assessment**. The starting point is the division of the bank's organization into business lines or units in order to assess and get better view of all topics. There are few different forms of self-assessment. One of them is the *checklists* under which the responsible personnel for one business line answer the list of the questions. The questionnaires should provide the insight into risk profile for the corresponding business line. Obviously, one of the disadvantages of the self-assessment is its highly subjective result. To avoid this usually more than one person do the checklists and later the results are independently verified.

The other form of self-assessment is the *workshop* and, generally, it is facilitated by some experts i.e. independent person. During the workshop the responsible personnel discuss about risk exposure, controls and what requirement actions should be done for improvement. Similarly to the checklists, the workshop is done for every business unit.

Usually, risk self-assessment is used as the first step for those firms which begin to assess their risk exposure. Often the internal database of loss events is still not collected and the self-assessment is the only approach that can give the basic view of the risk. Since the self-assessment is a forward looking approach it can also forecast risk frequency and severity for corresponding business line and event type. Its advantage is not only viewed when the internal database does not exist, but also when the lots of changes of business conditions or significant changes in size of business leads to useless of the old historical data. As already mentioned the self-assessment is subjective approach and it is necessary to validate the result using internal database, external database and/or by opinions of excepts. Nevertheless, the best way for assessing the risk exposure is the combination of internal database and risk self-assessment.

Measuring. After the risk assessment and control framework are conducted firm needs to be familiar, also, with quantitative measure of the risk exposure. That measure should provide more information on the result of control process and the changes in risk profile during the time. Nowadays, different types of measure are in use. We will consider the following ones: Key Risk Drivers (KRDs), Key Risk Indicators (KRIs), Loss Historical Data, Causal Models and Capital Models.

Risk type	KRD	KRI
1. Internal Fraud	1. Management & Supervision Recruitment Policy Pay Structure	1. Time Stamp Delays (Front Running)
2. External Fraud	2. Systems Quality	2. Number of Unauthorized Credit Card Transaction
3. Employment Practices & Workplace Safety	3. Recruitment Policy (Discrimination) Pay Structure Safety Measures	3. Number of Employee Complaints Staff Turnover Time Off Work
4. Clients, Products & Business Practices	4. Product Complexity Training of Sales Staff	4. Number of Clients Complaints Fines for Improper Practices
5. Damage to Physical Assets	5. Location of Buildings	5. Insurance Premiums
6. Business Disruption & System Failures	6. Systems Quality Back up Policies Business Continuity Plans	6. System Downtime
7. Execution, Delivery & Process Management	7.Management & Supervision Recruitment Policy (Qualification) Volume of Transactions Training of Back Office Staff	7. Number of Failed Trades Settlement Delay Errors in Transactions Processing

Table 2.4: Some KRDs and KRIs for Different Risk Event Types

Key Risk Drivers. The inherent risk profile of the organization is the risk profile when no control process is performed. The goal of obtaining this measure is to get an insight of the risk exposure before taking any control step and comparing it after the control is done. Key Risk Drivers are those drivers (factors) that give the measure of the inherent risk profile. Namely, they reflect the business environment and internal control systems. If the measure of the KRDs is changed than, also, the risk profile is changed. Table 2.4 presents examples of KRDs and KRIs for different risk event types. Important fact about KRDs is that they provide a forward-looking on the risk profile, and, therefore can be used in forecasting.

Key Risk Indicators. KRIs are the measures, based on data, that indicate the risk profile of the particular business unit or activity. The examples of the types of KRIs are listed in Table 2.4. Indicators should be easily quantify measure since they are, often, measured daily, and they should be risk-sensitive. KRIs have their trigger levels (also called escalation criteria, thresholds) which are designed to warn management when the acceptable risk level is exceeded. The selection of the threshold is different for every indicator. Generally, green, yellow and red threshold are associated with suitable quantitative measure of the indicator. Green level corresponds to the properly controlled risk, yellow means that risk is approaching to the unacceptable level, while the red signifies that risk has exceeded the acceptable level. KRIs, like KRDs, give a forward-looking and therefore they can be used in causal models. Moreover, these tools are also used in risk assessing and monitoring processes.

Loss Historical Data. Every bank organization should have their own internal database of OR losses. Without internal database bank could not be able to use the Advanced Measurement (AMA) capital models proposed by Basel II. Particulary, under AMA models bank are required to collect at least 3 years of loss experience (up to 5 years). Obviously, with own internal database it is possible to get insight of the risk exposure for every business line (BL) and event type (ET). It is easier to find out how frequent some types of risk are or how severe some operational risks can be. In addition, the internal database helps a lot in identification of risk types and control framework.

Causal Model. These models are the mathematical approaches to the forecasting of potential operational risks. They include: multifactor models, Bayesian or causal networks, fuzzy logic and neural networks. These models use KRDs, KRIs, internal loss database, external database in order to get multivariate distribution of losses. The aim is to find out which factors or factor have the major impact on a particular risk. The change of the model's factors should predict the risk exposure. However, since there is a lack of data considering OR losses, the limitation of the causal models is its requirement of many data points.

Capital Model. Capital models are the models that calculate the economic capital and regulatory capital charge. Basically, to determine the capital charge one should quantify the unexpected losses. As already mentioned Basel II has defined three basic models, that is Basic Indicator Approach (BIA), Standardized Approach (SA) and Advanced Measurement Approaches (AMA) methods. The discussion about these models will be done in the following chapter.

Monitoring. Sometimes it is difficult to separate which tools and models are used for measurement and which for the risk monitoring. It is often the case that these processes go together and that they use the same tools. Therefore, KRDs and KRIs are very important for the monitoring of OR. They can help management to understand the risk profile, potential changes in risk exposure and make attention to risks which are at unacceptable level. During the monitoring process the management should consider all risk measures, analyze the trends and gaps in the OR management framework. Further, the results of the risk assessment should be considered, and also the responsibilities in the processes together with appropriateness of action plan.

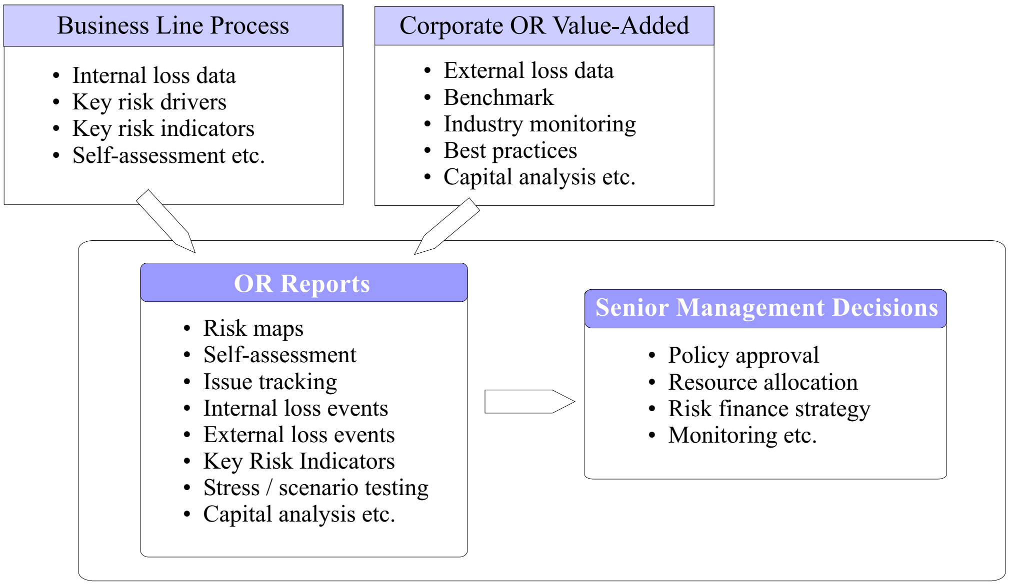

Figure 2.4: The OR Management Reporting Process

Reporting. Reporting is the final stage of OR process. In Figure 2.4 the OR management reporting is described. There are two parts of reporting. The first part is the business lines' reporting on the basis of internal loss data, KRDs, KRIs and sell-assessment. The second part is corporate OR reporting which tends to add-value by using external loss data, capital analysis, industry monitoring, best practices etc. The reports should be acceptable for the business managers and should also satisfy the senior management. After getting all reports in one period of time the senior management is required to create the overall risk profile, finance strategy and to approve the action plan.

Chapter 3

The Capital Models

3.1 The Review of Capital Models

Dealing with OR's measures is a relatively new field of research. It could be said that just after the Basel II has permitted a substantial degree of flexibility within the advanced models, the interest in calculating the OR capital charge has increased. This is on the account of the fact that these models leave the possibility for the bank to calculate the capital charge according to their own internal methods which should be based on statistics and mathematics theory.

In this section the basics of three Basel's capital models will be discussed together with the emphasis on the Loss Distribution Approach. As mentioned, Basel II distinguishes three models for calculating the OR capital charge: the Basic Indicator Approach (BIA), the Standardized Approach (SA), and the Advanced Measurement Approaches (AMA).

3.1.1 The Basic Indicator Approach

The BIA calculates the required capital for OR as a fixed percentage α of positive financial indicator averaged for the previous three years. Usually, the financial indicator is the annual gross income. Let us denote the k-th year financial indicator with FI_k. Then, the capital charge CC may be expressed as

$$\mathrm{CC}_{BIA} = \frac{\sum_{k=1}^{3} \max(\alpha \cdot \mathrm{FI}_k, 0)}{3} \tag{3.1}$$

where α is currently set at 15%. [1]

This is the most simple capital model and its use is recommended only for small-sized banks and/or for banks which begin the implementation of OR management and measurement. It is important to note that under this approach no qualifying criteria, listed in section 2.3, for management of OR are needed. Obviously, the basic limitation of this capital measure is no risk-sensitivity, and hence, lack of information on the actual risk exposure.

3.1.2 The Standardized Approach

The SA is more risk-sensitive method for calculation of OR required capital charge, which extends the BIA by decomposing banks activities into eight business lines (BL). Within each BL a financial indicator (e.g. annual gross income, asset size of BL) is multiplied by appropriate fixed percentage β. If we keep the notation, FI_{ki} will be k-th year financial indicator for the i-th BL, while β_i will be a corresponding percentage. Then, the total capital charge is just the average of the three-year regulatory capital charges summed across BLs, and can be expressed as

$$\mathrm{CC}_{SA} = \frac{\sum_{k=1}^{3} \max\left(\sum_{i=1}^{8} \beta_i \cdot \mathrm{FI}_{ki}, 0\right)}{3} \tag{3.2}$$

where β_i varies from 12% to 18%. [2]

Even thought SA is an improvement of BIA it does not reveal too many information on the insight of OR. Namely, it tells in which BL the risk exposure might be bigger, but it is not precise enough. Consequently, BCBS has developed the new AMA capital model.

3.1.3 The Advanced Measurement Approaches

The AMA are the most sophisticated capital models that allow bank to hold regulatory capital for OR based on its own internal models. It means that banks have the opportunity to develop their own procedures for measuring and assessing their exposure to OR. Dispute this given freedom, the usage of the AMA is subject of supervisory approval, and also, under this approach banks are, required to

[1] At first, α was set at 20%, later was revised at 12%, while at the moment it is 15%.

[2] Concretely, the values of β for every business line are: Corporate Finance 18%, Trading and Sales 18%, Retail Banking 12%, Commercial Banking 15%, Payment and Settlement 18%, Agency Services 15%, Asset Management 12% and Retail Brokerage 12%.

adopt some qualitative and quantitative criteria set by the BCBS, listed in section 2.3.

Under AMA the total capital charge is, simply, the sum of the figures of **expected loss** (EL) and **unexpected loss** (UL).

$$\text{CC}_{AMA} = Expected\ Loss + Unexpected\ Loss$$

What is not simple, is to obtain these figures. According to Basel's rules, bank must be able to demonstrate to supervisors that the risk measure, used for regulatory capital purposes, reflects a holding period of **one year** and a **confidence level of 99.9 percent**. In fact, banks' internal models should capture potentially severe tail loss events, i.e. models should be able to produce reasonable estimates of unexpected losses.

The way of calculation of capital charge highly depends on information which banks have and use. Namely, there are four groups of information: internal data, external data, scenario analysis and factors (i.e. KRD's, KRI's). Following this, within AMA approaches four basic models can be distinguished:

- The Loss Distribution Approach (LDA),
- Internal Measurement Approach (IMA),
- The Scorecard or Risk Drivers and Control Approach, and
- The Scenario-based Approach.

The LDA and IMA are mostly based on the usage of internal loss data, the Scorecard approach mainly considers the KRD's and KRI's, while, the Scenario-based approach uses various scenarios to evaluate bank's risk. However, in practice most banks uses elements of all four approaches.

Our study is mainly concentrated on LDA model as it is founded on statistical and actuarial theory and gives a possibility of implementation of various mathematical models in order to derive the most appropriate figure for bank's OR capital charge. Next section describes the standard LDA and gives a mathematical formulation of the underlying model.

3.1.4 Loss Distribution Approach

The LDA is founded on standard actuarial theory which considers frequency of losses and severity of losses as independent random variables. From the economic

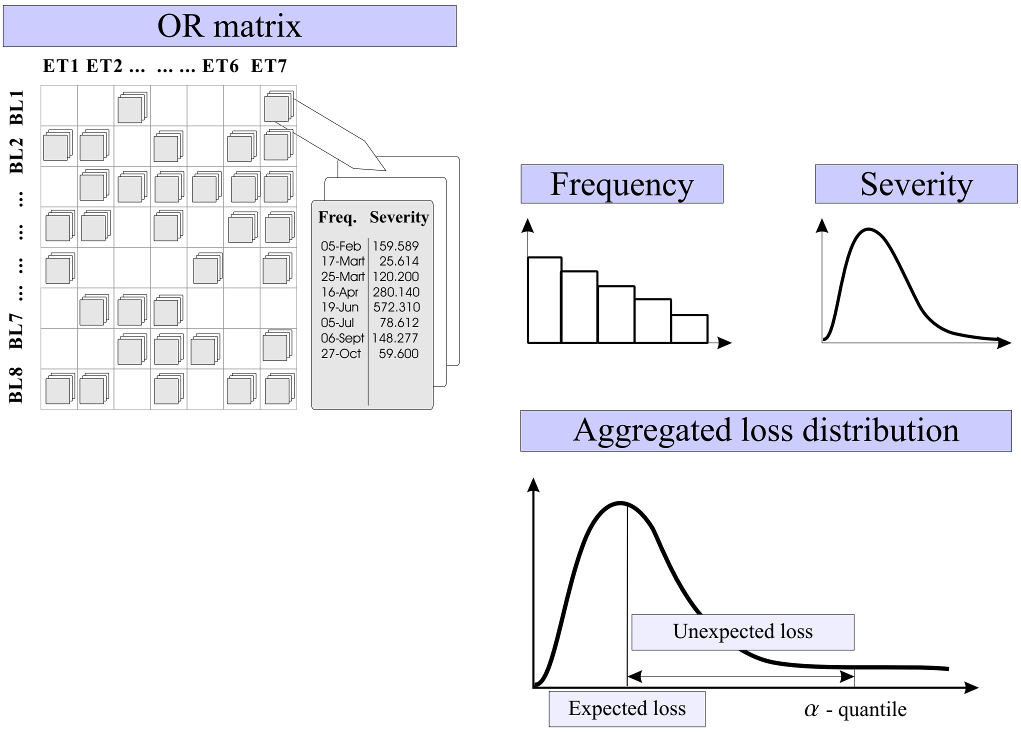

Figure 3.1: LDA

point of view, this division of the operational losses has justification in the fact that some of management's actions would affect only a severity of losses and some just a frequency. Therefore, the frequency and the severity will be fitted with different distribution functions, and the **aggregated loss distribution** will be an analytical (if it is possible) or simulated form of both chosen distribution functions. The aim of this approach is to compute the aggregated loss distributions from which the total OR capital charge can be derived.

Certainly, the LDA can have different forms as every organization can modified it according to its own needs. However, there are some basic steps in its implementation which need to be performed.

Basically, LDA uses a bank's internal loss data. According to the Basel II, banks are required to collect at least **three years of history loss data** in order to adopt the LDA. These internal loss data need to be arranged into a 56-cell OR matrix (8 BL $\times$ 7 ET). Each element in the matrix is defined by its business line (BL), where it has occurred, and event type (ET), what type of loss has occurred. See Figure 3.1.

The standard LDA consists of the following 4 steps:

1. For every BL-ET combination frequency and severity should be fitted with appropriate distribution functions. The parameters for these distributions can be estimated via Method of Maximum Likelihood Estimation (MLE), Method of Moments or some other such as Expectation-Maximization algorithm (EM) etc.

2. The aggregated distribution is a compounded distribution of the selected frequency and the severity distributions. Usually, its analytical form is difficult to get and hence, instead, the simulation techniques are applied (e.g. Monte Carlo simulation).

3. The figure of OR capital charge is calculated for every BL-ET combination as a sum of expected loss and unexpected loss of corresponding aggregated distribution.

4. Finally, the total OR capital charge is just the sum of all calculated OR capital charges across the matrix, if the correlation of loss events is assumed to be one.

Suppose that we are given the bank's OR internal loss data for some period of time $\Delta t = T_2 - T_1$. If X denotes a random variable that represents bank's loss event in time interval Δt, then n-times realization of X is our given sample i.e. random vector $\mathbf{x} = (x_1, \ldots, x_n)$. Obviously, X is from $\mathbb{R}^+$ since we are dealing with amount of losses.

The observed data sample should be classified into OR matrix. Let i and j be indices that denote given BL and ET, respectively. The particular (i, j) cell for BL_i-ET_j combination will be observed as a separate class for which the following notation is defined.

- X_{ij} is a random variable for the **severity of losses** in (i, j) cell and its probability function is denoted as $f_{ij}(x)$ and cumulative distribution function as $F_{ij}(x) = \mathrm{P}(X_{ij} \leq x)$.

- N_{ij} is a random variable of the **frequency of losses** in (i, j) cell. Its corresponding probability function is $p_{ij}(k) = \mathrm{P}(N_{ij} = k)$ and cumulative distribution function $P_{ij}(n) = \mathrm{P}(N_{ij} \leq n)$ i.e.

$$P_{ij}(n) = \sum_{k=0}^{n} p_{i,j}(k) \ .$$

- Following notation, the total (cumulative) loss L_{ij} for (i, j) cell in time interval $[T_1, T_2]$ is

$$L_{ij} = \sum_{k=1}^{N_{ij}} X_{ij}(k) = X_{ij}(1) + X_{ij}(2) + \cdots + X_{ij}(N_{ij}) \, , \tag{3.3}$$

where $X_{ij}(k)$ is an amount of k-th loss occurred in (i, j) cell. The cumulative distribution function for random variable L_{ij} is called the **aggregated loss distribution**, here in notation G_{ij}. Its analytical form is a compound distribution given by

$$G_{ij}(x) = \mathrm{P}(L_{ij} \leq x) = \begin{cases} \sum\limits_{n=1}^{\infty} p_{ij}(n) F_{ij}^{n\star}(x) & , \quad x > 0 \\ \\ p_{ij}(0) & , \quad x = 0 \end{cases} \tag{3.4}$$

where $n\star$ denotes n-fold convolution on distribution functions F with itself.

Certainly, in order to simplify the model and to follow the Basel II recommendations we need to put some standard assumption on the properties of random variables N and X within every cell.

Assumption 1. The frequency N_{ij} and the severity of losses X_{ij} are independent random variables.

Assumption 2. The severity X_{ij} is an independent and identically distributed (i.i.d.) random variable.

From the first assumption it is clear that frequency and severity are treated as two independent sources of randomness. By this assumption the possibility of correlation among frequency and severity within one cell is completely rejected. The second assumption means that two different losses within the same cell are independent and identically distributed. This allows us to consider one cell as separate class.

If we suppose that for every cell the G_{ij} distribution is obtained then the estimation of **OR capital charge** can be performed in the following way. Firstly, we will calculated it at the level of one cell and later for the whole matrix. The OR capital charge is a sum of corresponding expected and unexpected loss within (i, j) cell, in notation

$$\mathrm{CC}_{ij} = \mathrm{EL}_{ij} + \mathrm{UL}_{ij} \, . \tag{3.5}$$

The EL for some particular cell is usually defined as a mean value of a corresponding total loss random variable i.e.

$$\begin{aligned} \mathrm{EL}_{ij} &= \mathrm{E}(L_{ij}) \\ &= \int_0^\infty x \cdot \mathrm{d}G_{ij}(x) \; . \end{aligned} \tag{3.6}$$

Yet, some other measures for EL can be considered, such as median value. If the loss sample has a high skewness and kurtosis properties with big number of outliers then median measure can be more suitable for calculation of expected loss.

On the other side, the UL figure should capture the tail of the distribution, and hence, it is computed as the difference between distribution's α-th quantile and expected loss. The α-th quantile of distribution can be thought of as **Value at Risk measure** (VaR). Namely, VaR at confidence level α for a period of time Δt, in notation $\mathrm{VaR}_{\alpha,\Delta t}$, is the smallest loss that is greater than the α-th quantile of some given distribution F_Y, i.e.

$$\mathrm{VaR}_{\alpha,\Delta t} \; := \; \inf \{ y : F_Y(y) \geq \alpha \} \; . \tag{3.7}$$

In other words, $\mathrm{VaR}_{\alpha,\Delta t}$ presents the expected maximum loss over the time interval within a given confidence level, and it is a result of

$$\mathrm{P}(Y_{t+\Delta t} - Y_t > \mathrm{VaR}_{\alpha,\Delta t}) = 1 - \alpha \; .$$

Accordingly, the UL can be defined as

$$\begin{aligned} \mathrm{UL}_{ij,\alpha} &= \inf \{ x : G_{ij}(x) \geq \alpha \} - \mathrm{E}(L_{ij}) \\ &= \mathrm{VaR}^{ij}_{\alpha,\Delta t} - \mathrm{E}(L_{ij}) \; . \end{aligned} \tag{3.8}$$

From Equation 3.6 and 3.8 it follows that the capital charge in (i, j) cell for a confidence level α in time interval Δt is equal to VaR measure, that is

$$\mathrm{CC}_{ij,\alpha} = \inf \{ x : G_{ij}(x) \geq \alpha \} = \mathrm{VaR}^{ij}_{\alpha,\Delta t} \; . \tag{3.9}$$

In context of Basel II, banks are required to set α confidence level at 99.9 % and a time interval Δt to 1 year.

Now, we can obtain the estimate of **total capital charge** i.e. the capital charge for the whole OR matrix. Usually, in practice the total capital charge is simply expressed as a sum of all calculated capital charges across the matrix i.e.

$$\begin{aligned} \mathrm{CC}_\alpha &= \sum_i \sum_j \mathrm{CC}_{ij,\alpha} \\ &= \sum_i \sum_j \mathrm{VaR}^{ij}_{\alpha,\Delta t} \,. \end{aligned} \tag{3.10}$$

Indeed, this is true only if a perfect correlation among aggregated losses L_{ij} is assumed.

Therefore, if we assume that the aggregated losses are correlated with some other degree the formula (3.10) need to be modified. The section 4.7 deals with these issues and gives other expression of total OR capital charge.

Chapter 4

Modelling OR Losses Under LDA

In this chapter the LDA approach is given in more details. We have followed the idea of what an individual bank should do in order to implement this approach, and also, which mathematical models and techniques should be used.

The chapter is organized as follows. Firstly, in Section 4.1, we start with descriptive statistic analysis of internal data, and then, in Section 4.2 we moved to its parametrical fitting. Both the frequency and the severity of loss data are considered. The most common distribution functions are listed and reviewed. Secondly, in Section 4.3 the parameters estimation is presented through standard and adjusted approach. The first approach treats OR internal data set as complete and uses Maximum Likelihood parameters estimation method. The second approach considers the OR data as truncated and, therefore, uses adjustment of parameters. Moreover, the issue of parameters estimation for incomplete data is explained via Expectation-Maximization algorithm. Thirdly, in Section 4.4. the Kolmogorov-Smirnov statistical test for the goodness of the fit is applied to the chosen distribution functions. Also, the adjusted test for left-truncated distributions is presented. Further, in Section 4.5 the problem of fitting the tail of severity distribution is explained through the Extreme Value Theory. And finally, in Section 4.6 the Monte Carlo method for simulation of aggregated loss distribution is given, and in Section 4.7 the correlation effect of loss events is considered and the required total capital charge figure is formulated.

4.1 Descriptive Statistic Analysis

Performing the descriptive statistic analysis is the first step in modelling the internal loss data. Obviously, we need to get familiar with underlying structure of the

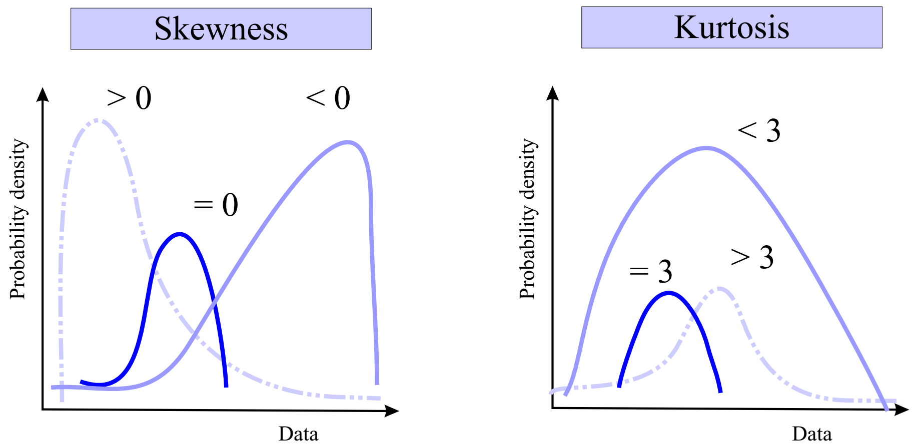

Figure 4.1: Skewness and kurtosis

available data through its moments, empirical distributions, graphical inspection etc.

Descriptive Statistics. The descriptive statistic provides information on number of observed data, its minimum and maximum value, the mean value, the standard deviation, the skewness, the kurtosis etc. The most important measures in OR loss modelling are skewness and kurtosis.

The skewness is the third standardized moment which explains the asymmetry of data around the mean value, while kurtosis is the forth standardized moment for measure of tail-heaviness. Therefore, these measures can be considered as guidelines in choosing the distribution for fitting the data.

Namely, for the normal distribution skewness is equal to zero. See Figure 4.1. Positive skewness means that distributions have more probability mass concentrated on the left side of the mean, while, negatively skewed distributions have on the right side. The kurtosis of normal distribution is equal to 3 and distributions with more heavier tails than the normal distribution, have kurtosis greater than 3. Respectively, distributions with ticker tails have kurtosis less than 3.

Graphics. In our empirical analysis various plots are adopted in order to provide information about visual characteristics of data.

Firstly, the box-plot provides information about sample percentiles and outliers. See Figure 4.2. In the box the lower quartile (25^{th} quantile), median (50^{th} quantile), and upper quartile (75^{th} quantile) values are shown, while lines at the end of the box show spread of sample. Obviously, outliers are data beyond the box and lines.

Secondly, the histogram gives information about frequency of data. See Figure 4.2.

Thirdly, information on data's probability structure can be obtained using the kernel smoothing method. This technique describes data by estimating its density in a nonparametric way. That is, every data point from the random vector $\mathbf{x} = (x_1, \ldots, x_n)$ is included through the kernel density estimator

$$\hat{f}(x,h) = \frac{1}{nh}\sum_{k=1}^{n} \mathrm{K}\Big(\frac{x - x_k}{h}\Big)$$

where h is a scaling factor (bandwidth) and K chosen kernel function. The Epanechnikov kernel is the most common one and it is defined as

$$\mathrm{K}(\eta) = \begin{cases} \frac{3}{4}(1-\eta^2) & , \quad \eta \in (-1,1) \\ 0 & , \quad \text{else} \end{cases}$$

where $\eta = (x - x_k)/h$.

The result is the graphic of kernel smoothing estimator that gives an empirical version of a probability density function as shown in Figure 4.2. Namely, instead of using a parametric density function and estimating the parameters, we produce a nonparametric density estimate that tries to fit the data.

Obviously, the data's estimated probability function should be supplemented by an empirical cumulative distribution function (cdf). It is the function that assigns probability for a sample $\mathbf{x} = (x_1, \ldots, x_n)$ as a number of observations less or equal to x divided by the size of sample n, i.e.

$$F_{ecdf}(x) = \frac{1}{n}\sum_{k=1}^{n} \mathrm{I}(x_k \leq x) \tag{4.1}$$

where I is a identical function

$$\mathrm{I}(x_k \leq x) = \begin{cases} 1 & , \quad x_k \leq x \\ 0 & , \quad x_k > x \end{cases} \tag{4.2}$$

In Figure 4.2 the empirical cdf is presented.

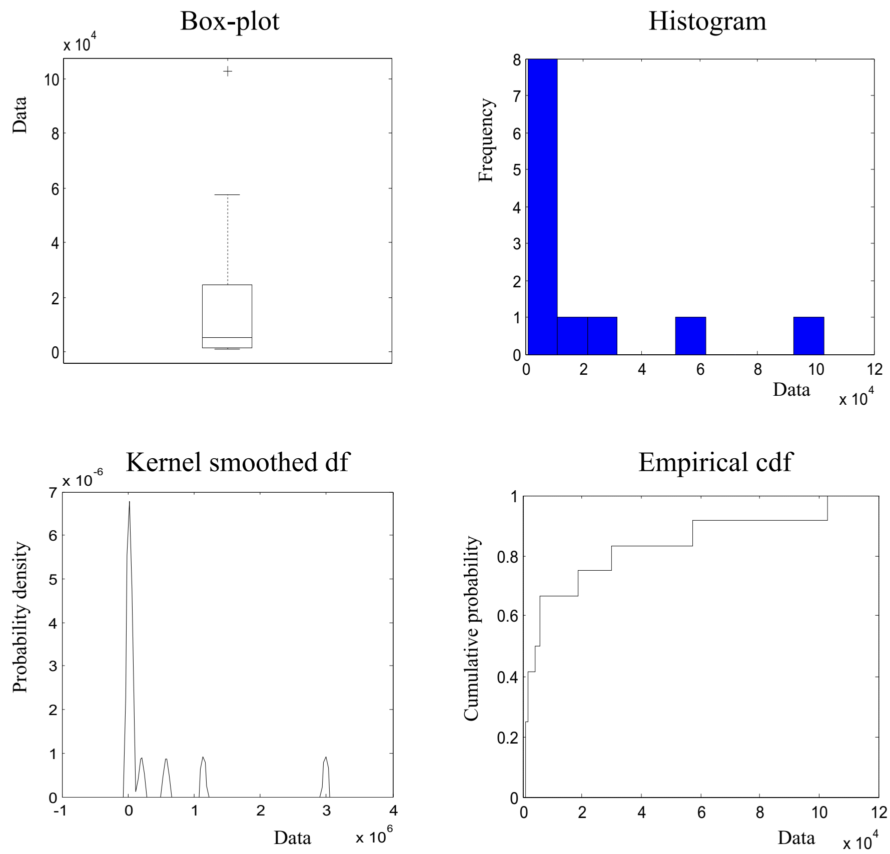

Figure 4.2: Boxplot, Histogram, Epanechnikov pdf, Empirical cdf

Tail Plot. Modelling the severity random variable assigns problems considering the tail of the distribution. It fact, more attention has to be put on tail thickness in order to explain the highest severity OR loss data. One of the common tools for graphical presentation of the tail is the tail plot.

If we find the empirical cdf F_{ecdf} for random sample $\mathbf{x} = (x_1, \ldots, x_n)$, then, $\log(1 - F_{ecdf}(x))$ plotted on the vertical axis against $\log(x)$ presents the tail plot. See Figure 4.3. From the linear shape of the plot it can be concluded that data are probably drawn from heavier-tailed distribution (e.g. Pareto-type distribution). However, a preliminary estimate of the tail parameter is a slope a from the line

$$\log(1 - F_{ecdf}(x)) = -a \cdot \log(x) + b \qquad (4.3)$$

where b is a constant. The line (4.3) of the slope -1 is considered as a reference line, since if the plot is near or above it, the tail parameter is considered equal or greater then 1, meaning the data have heavy tails. These issues were investigated in more details by P. de Fontnouvelle and E. Rosengren in paper [6].

Mean Excess Plot. Another useful graphical instrument for better understanding of distribution's tails is the mean excess plot. It is based on mean excess function $\mathrm{MEF}(u)$ which is defined for loss sample $\mathbf{x} = (x_1, \ldots, x_n)$ as conditional expectation of number of losses exceeding the chosen threshold u given that the losses are bigger then threshold. That is,

$$\mathrm{MEF}(u) = \mathrm{E}(X - u \mid X > u) \ .$$

Its estimate $\widehat{\mathrm{MEF}}(u)$ is defined as average of all excesses over threshold u minus the threshold itself

$$\widehat{\mathrm{MEF}}(u) = \frac{\mathrm{I}(x_k > u) \sum_{k \leq n}(x_k - u)}{\sum_{k \leq n} \mathrm{I}(x_k > u)} \ .$$

Similar to the tail plot, the shape of the mean excess plot gives an information on the underlying distribution. In Embrechts, [9] it is stated that if the plot shows a upward trend (positive slope) it implies that the data belongs to heavier tailed distribution. If the plot is more horizontal line then the data are exponentially distributed, while downward trend (negative slope) is sign of light-tailed distribution.

Certainly, having results of descriptive statistical analysis simplify the process of finding which distribution should be used for fitting frequency N and severity X random variables. All this measures and plots are performed in our empirical study and the results are presented in Chapter 5.

4.2 Parametrical Fitting

In this section different parametric distribution functions are presented. Some of them are used in fitting frequency random variable N and some for severity random variable X. Their brief review is given.

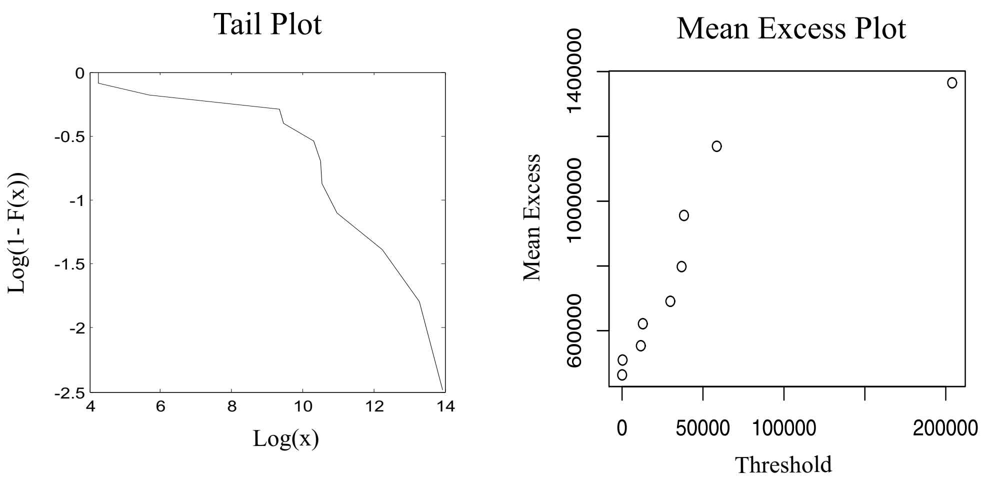

Figure 4.3: Tail Plot and Mean Excess Plot

4.2.1 Frequency

Considering a frequency, the most common and the most appropriate distributions are Poisson and Negative Binomial. Both are discrete parametric distributions that can well explained the counting problem i.e. the number of occurrence of a random event in a given time interval. Their plots are shown in Figure 4.4.

The **Poisson distribution** with its simple form and properties is a good candidate for the start of modelling the frequency of OR loss events. Its probability mass function is defined as

$$f(n;\lambda) = \frac{\lambda^n}{n!} e^{-\lambda} \ . \tag{4.4}$$

It takes nonnegative integer values and has one parameter λ (intensity rate) which is, also a value of a mean and a variance of the distribution.

However, **Negative Binomial distribution** can be, sometimes, more suitable than Poisson distribution since it has a variance greater than its mean. This property gives more variability for the interval of the expected number of events. The Negative Binomial distribution has two real parameters p and r, $p \in (0,1)$, $r > 0$, and probability mass function defined for $n \in \mathbb{N} \cup \{0\}$ as

$$f(n;r,\ p) = \binom{r+n-1}{n} p^r (1-p)^n \ . \tag{4.5}$$

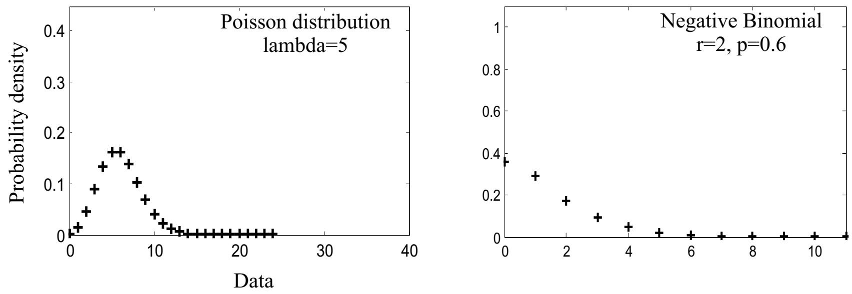

Figure 4.4: Plot of Poisson and Negative Binomial Distribution

If we consider the frequency of losses in non discrete way, then an arrival process can be introduced. Such process is obviously, irregularly arranged in time, since there is no rule when the next loss event will happen. Consequently, we need assumption that the **loss events are independently distributed**.

In general, this kind of process is modelled by stochastic **Poisson Process** N_t which has the following properties:

1. $N_0 = 0$.

2. N_t has independent increments. That is, for every $n \in \mathbb{N}$ and any $t_1, t_2, \ldots, t_n$ such that $0 \leq t_1 \leq t_2 \leq \cdots < \infty$ the random variables $N_{t_{i+1}} - N_{t_i}$, for $i = 0, 1, \ldots, n$ are independent.

3. N_t is homogeneous. That is, for any $s, t, \Delta t > 0$ random variables $N_{s+\Delta t} - N_s$ and $N_{t+\Delta t} - N_t$ are identically distributed.

4. The number of observations n in an interval Δt has Poisson distribution with intensity rate $\lambda \Delta t$, $\lambda > 0$. That is, for every $t, \Delta t > 0$

$$\mathrm{P}(N_{t+\Delta t} - N_t = n) = \frac{(\lambda \Delta t)^n e^{-\lambda \Delta t}}{n!} \ .$$

There are two types of Poisson process: a homogeneous (HPP) and a non-homogeneous (NHPP). The HPP is the ordinary Poisson distribution with intensity rate λ constant over time as defined by function (4.4). In contrast to HPP, NHPP has intensity rate which changes during the time and is defined by deterministic intensity function $\lambda(t)$. In general, a cumulative intensity over a given time interval $\Delta t = T_2 - T_1$ is equal to

$$\hat{\lambda} = \int_{T_1}^{T_2} \lambda(t) \mathrm{d}t \ .$$

The type of function $\lambda(t)$ highly depends on the particular data which are fitted. However, in the literature (see A. Chernobai 2005, [13], R. Giacometti 2007, [17]) two types are commonly used in OR frequency modelling: LogNormal cdf-like and LogWeibull cdf-like cumulative intensity. They are defined as

$$\begin{aligned} \text{LogNormal cdf-like} \quad \hat{\lambda}(t) &= a + \frac{b\exp(\frac{-\log^2(t-d)}{2c^2})}{\sqrt{2\pi c}} \\ \text{LogWeibull cdf-like} \quad \hat{\lambda}(t) &= a - b \cdot \exp(-c\log^d(t)) \end{aligned}$$

where a, b, c, d are parameters.

In our empirical study Poisson and Negative Binomial distribution functions are considered.

4.2.2 Severity

The descriptive statistical analysis of severity of OR loss data, in most situations, shows a positive skewness and a high kurtosis. Therefore, for fitting a severity random variable X we should consider distributions with heavier tails. Indeed, we have to be aware of the fact that the size of losses in internal data can be small/medium and very large. Namely, small/medium-sized losses are generated by high-frequency low-impact events and these losses constitute the body of severity distribution. The second one, the large losses are generated by low-frequency high-impact events and constitute the tail of distribution. Consequently, if our data set consists of more severe losses then more heavier tailed distributions should be considered for its fitting.

According to Basel Committee's surveys conducted by Risk Management Group in 2002 and in 2004, the most commonly used distribution for severity of data is **LogNormal**. However, there is quite a number of distributions that can be adopted in modelling severity of OR losses like Exponential, LogNormal, LogLogistic, Weibull, LogWeibull, Pareto, General Pareto Distribution (GPD), Burr, Gamma, LogGamma, log-α stable etc. Certainly, the nature of a given data determines which distribution gives better fit.

In our empirical study Weibull, LogNormal, LogLogistic and Pareto distribution functions are considered[1]. If we classify the distributions by their tail heaviness then Weibull distribution comes from light tailed, LogNormal and LogLogistic distributions form medium tailed and Pareto from heavier tailed distributions.

[1] We have also considered Exponential distribution function, but since the fit of severity of every data set did not provide good results, the consideration of this distribution is omitted from the paper.

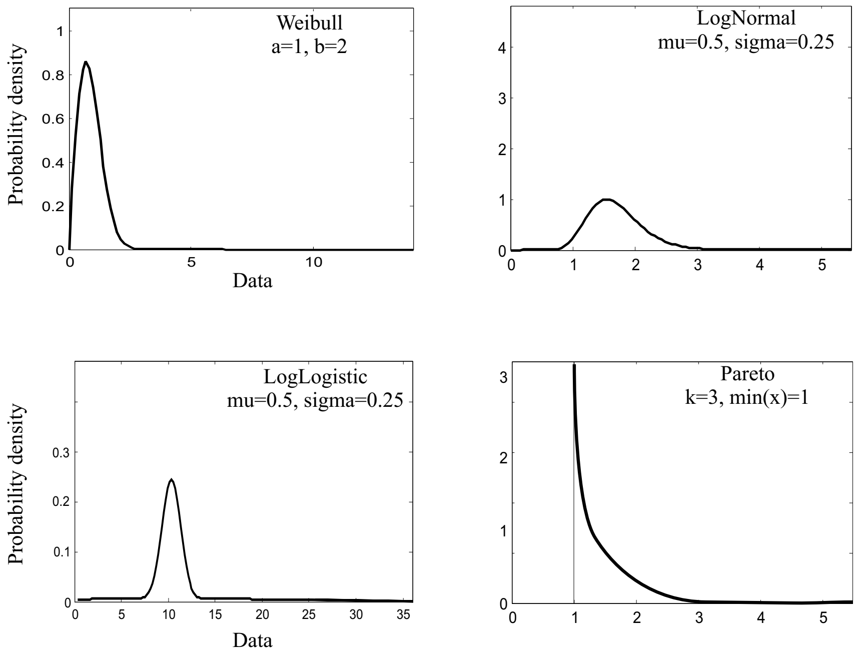

Figure 4.5: Distribution Plots

Theirs probability density functions are defined as following and plots shown in Figure 4.5.

$$\text{Weibull} \quad f(x;a,b) = \frac{a}{b}\left(\frac{x}{a}\right)^{b-1} \exp\left(-\left(\frac{x}{a}\right)^{b}\right), \quad x > 0, \;\; a,b > 0$$

$$\text{LogNormal} \quad f(x;\mu,\sigma) = \frac{1}{x \cdot \sigma\sqrt{2\pi}} \exp\left(\frac{-(\ln x - \mu)^2}{2\sigma^2}\right), \quad x > 0$$

$$\text{LogLogistic} \quad f(x;\mu,\sigma) = \frac{\exp(z)}{x \cdot \sigma(1+\exp(z))^2}, \quad z = \frac{\ln x - \mu}{\sigma}, \quad x > 0$$

$$\text{Pareto} \quad f(x;k,x_m) = \frac{k \cdot x_m^k}{x^{k+1}}, \quad x \in [x_m, +\infty), \;\; k > 0$$

4.3 Parameters Estimation

One of the most important thing in distributions' parameters estimation is the knowledge of what kind of data set we have. The standard approaches deal with complete data set, while some others are modified in order to capture the effect of missing data.

It was said before that banks only recently have started to collect information on theirs OR loss events, and thus, most of internal database are not complete. This is on account of the fact that most banks are not able to record all OR losses, especially at the beginning of theirs OR management process. Instead, they specified some threshold u, below which OR losses will not be fully recorded. The incompleteness can be viewed through censorship or truncation of data. If only frequency, below u, is not recorded then data are considered **censored** from below, and if both frequency and severity of losses are not known then data are said to be **truncated** from below.

Particulary, OR loss data can be either left-censored or left-truncated, and in most cases, they are left-truncated. This specific property of OR loss data can be (and should be) included in parameters estimation of distribution functions of both the frequency and severity. Recently, there are more and more papers that investigate these issues such as N. Baud 2002, [11], A. Chernobai 2005, [13], A. Chernobai 2006, [15], R. Giacometti 2007, [17]. The reason is simple. The obtained sample fitted distributions under standard parameter estimation method are in fact different from the true loss distributions. We need to use conditional distributions, that is the probability distributions conditionally to losses greater than some specific threshold.

In this section two approaches will be given, the standard also called naive approach that ignore the effect of OR missing data, and adjusted approach that considers the exact nature of a given OR data.

1. **Standard Approach.** Standard approach is based on the full-data parameters estimation methods meaning that data are considered complete. The most common method is Maximum Likelihood Estimation (MLE) that consists of two basic steps. Firstly, the likelihood function for a given distribution is formed as an argument of a distribution parameter, and secondly it is maximized by that parameter.

   Particulary, let suppose that we have $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ i.i.d random vector drawn from X random variable with a density function $f(x; \theta)$, where θ is parameter or set of parameters ($\theta \in \Theta$). Then, the corresponding likelihood function is simply a product of density functions for that sample

$$\mathrm{L}_{\mathbf{x}}(\theta) = \prod_{k=1}^{n} f(x_k; \theta) \ . \tag{4.6}$$

$\mathrm{L}_{\mathbf{x}}(\theta)$ is a function of parameters θ when data $\mathbf{x}$ are fixed. What we want to find is the parameter $\hat{\theta}^{mle}$ that maximizes L function, that is

$$\hat{\theta}^{mle} = \operatorname*{argmax}_{\theta} \mathrm{L}_{\mathbf{x}}(\theta) \ . \tag{4.7}$$

Certainly, this maximization problem can be easy or hard to perform, since it highly depends on characteristics of distribution function f. Often likelihood function is replaced with log-likelihood function $\log \mathrm{L}_{\mathbf{x}}(\theta)$ in order to simplify the maximization. However, for simpler forms of distributions the maximum likelihood estimation is done by setting the first derivative of $\mathrm{L}_{\mathbf{x}}(\theta)$ to zero, and solving it directly for parameters θ.

2. **Adjusted Approach.** Under this approach the frequency and severity parameters are modified and adjusted by the missing information. We have considered left-truncated data for some threshold u.

   Obviously, for the given threshold u the true (complete) severity cumulative distribution function denoted by $F(x;\theta)$ gives information on probability if data falls above or below u. The first case when $x > u$ has a probability $\mathrm{P}(x > u) = 1 - F(u;\theta)$, and for $x \leq u$ the probability is simply $\mathrm{P}(x \leq u) = F(u;\theta)$. This means that in case of incomplete data, there is lack of information for the area $x \leq u$. Consequently, the parameters should be adjusted by the information from which area data come.

   Let us denote the adjusted severity distribution function by $f^{adj}(x;\theta)$, then f^{adj} is equal to conditional distribution given that loss x exceeded threshold u, i.e.

$$\begin{aligned} f^{adj}(x;\theta) &= f(x;\theta \mid x \geq u) && (4.8) \\ &= \mathrm{I}(x \geq u) \frac{f(x;\theta)}{\int_u^{+\infty} f(y;\theta)\mathrm{d}y} && (4.9) \\ &= \begin{cases} \dfrac{f(x;\theta)}{1 - F(u;\theta)} & , \quad x \geq u \\ 0 & , \quad x < u \end{cases} && (4.10) \end{aligned}$$

The above equations imply that complete distribution function is adjusted in terms of division by the probability that loss occurred is greater than threshold.

Now, from Equation 4.10 the new adjusted severity parameter should be computed. Following the MLE reasoning, the $\hat{\theta}^{adj}$ will be a maximum argument of a log-likelihood function for $f^{adj}(x;\theta)$, i.e.

$$\begin{aligned}
\hat{\theta}^{adj} &= \underset{\theta}{\operatorname{argmax}} \log \mathrm{L}_{\mathbf{x}}(\theta) \\
&= \underset{\theta}{\operatorname{argmax}} \log \Big(\prod_{k=1}^{n} f^{adj}(x_k;\theta) \Big) \\
&= \underset{\theta}{\operatorname{argmax}} \log \Big(\prod_{k=1}^{n} \frac{f(x_k;\theta)}{1-F(u;\theta)} \Big) \qquad (4.11) \\
&= \underset{\theta}{\operatorname{argmax}} \Big(\sum_{k=1}^{n} \log f(x_k;\theta) - n \log(1-F(u;\theta)) \Big) . \qquad (4.12)
\end{aligned}$$

If the severity parameter is computed then the frequency parameter should be also adjusted as

$$\hat{\tau}^{adj} = \frac{\hat{\tau}^{mle}}{1-F(u;\hat{\theta}^{adj})} \qquad (4.13)$$

where $\hat{\tau}^{mle}$ is MLE parameter estimate of frequency distribution $P(n;\tau)$.

Apparently, only problem in this approach can be in maximization of (4.12). Indeed, in some cases it is hard to obtain the parameter analytically, and hence, standard numerical optimization tools should be considered. It is evident that having computed severity parameter $\hat{\theta}^{adj}$ from Equation 4.12, the estimation of frequency parameter $\hat{\tau}^{adj}$ is straightforward from Equation 4.13.

Although, the maximization problem (4.12) can be solved numerically for almost every type of density distribution function, the other way of finding the adjusted parameters can be performed via **Expectation-Maximization** algorithm (EM). In the following sections the theoretical background of EM algorithm is given together with application to the most common Poisson-LogNormal aggregated loss distribution.

4.3.1 EM Algorithm - Theoretical Background

As mentioned before, the standard MLE method is formed for the estimation of distributions' parameters when data are assumed to be complete. In those cases when data are incomplete or has missing values, naturally, the MLE can not be fully acceptable. One of the iterative techniques that can be adopted in order to find the maximum likelihood parameter estimations for incomplete data set is EM algorithm.

The EM algorithm was described and explained by Dempster, Laird and Rubin in 1977, [8]. This paper has been a "trigger" for algorithm's wider application in various problems where missing data and unidentified variables are involved. Recently in risk measurement, specially in OR measurement, this EM algorithm has found its use.

As the name says EM has two steps: expectation (**E-step**) and maximization (**M-step**). The main goal is to compute an **expectation** of the likelihood function when the missing data are included, and then, to compute the parameters that **maximize** that expected likelihood function.

If we consider a left-truncated data for some threshold u then a set domain $A = [0, +\infty)$ is divided by that threshold into two sets, $A_1 = [0, u]$ and $A_2 = (u, +\infty)$. This means that a given data sample $\mathbf{x} = (x_1, \ldots, x_n)$ comes only from A_2 set, and thus, in A_1 set we do not know neither the quantity nor the severity of data. Consequently, we need to set some new notations.

Random variable X_1 will present the loss event from A_1 set, while random variable X_2 will present loss event occurred in A_2 set. In light of that, X_1 is considered as missing random variable and our given sample is now denoted as $\mathbf{x}_2 = (x_{21}, x_{22}, \ldots, x_{2n_2})$, $\mathbf{x}_2 \in (u, +\infty)^{n_2}$.

Now, the complete data sample is simply expansion of observed data $\mathbf{x}_2$ with random variable X_1, that is $\mathbf{x} = (X_1, \mathbf{x}_2)$. Obviously, $\mathbf{x}$ is a random draw from joint random variable $X = (X_1, X_2)$ where a random vector $\mathbf{x}_1 \in [0, u]^{n_1}$ is "missing" and replaced by X_1 random variable.

Consequently, the joint density function is, according to Bayes rule and the low of total probability, equal to

$$\begin{aligned} f_{X_1,X_2}(x_1, x_2; \theta) &= f(X_1, x_2; \theta) \\ &= f(X_1; x_2, \theta) \cdot f(x_2; \theta) \ . \end{aligned} \tag{4.14}$$

This means that complete data will be observed through the conditional function of missing data given the known data and the marginal function of known data. From Equation 4.6 the likelihood function for the joint likelihood function of complete data set is equal to

$$\begin{aligned}\mathrm{L}_{X_1,\mathbf{x}_2}(\theta) &= f(X_1,\mathbf{x}_2;\theta)\\ &= \prod_{k=1}^{n_2} f(X_1,x_{2k};\theta)\ . \end{aligned} \tag{4.15}$$

Note that now $\mathrm{L}_{X_1,\mathbf{x}_2}(\theta)$ is a random variable since it depends on X_1 and its maximization by parameter θ can not be performed via standard MLE model. This is the main reason why some other model must be considered. We have applied the EM algorithm. It is defined in the following way.

The EM Algorithm. Let a given data set $\mathbf{x}_2$ be expanded by a random variable X_1 of missing data. The joint random variable $X = (X_1, X_2)$ has joint density function $f(x_1, x_2; \theta)$ and likelihood function $\mathrm{L}_{X_1,\mathbf{x}_2}(\theta)$ defined by Equation 4.14 and Equation 4.15, respectively. Then, given an initial parameter $\theta^{(0)}$ the following steps are iterated.

E-step: Evaluate $Q(\theta, \theta^{(i)})$ as a conditional expectation of joint log-likelihood function

$$Q(\theta,\theta^{(i)}) := \mathrm{E}(\log \mathrm{L}_{X_1,\mathbf{x}_2}(\theta) \mid \mathbf{x}_2, \theta^{(i)})\ . \tag{4.16}$$

M-step: Find $\theta = \theta^{(i+1)}$ which maximizes $Q(\theta, \theta^{(i)})$ i.e.

$$\theta^{(i+1)} = \operatorname*{argmax}_{\theta} Q(\theta, \theta^{(i)})\ . \tag{4.17}$$

The basic idea of this iterative algorithm is to choose parameter $\theta^{(i+1)}$ that maximizes log-likelihood function $\log(\mathrm{L}_{X_1,\mathbf{x}_2}(\theta))$. But, since the analytical form of joint log-density function $\log f(X_1, \mathbf{x}_2; \theta)$ is not known, the EM algorithm maximize the current expectation of $\log(\mathrm{L}_{X_1,\mathbf{x}_2}(\theta))$ for given data $\mathbf{x}_2$ and current parameter $\theta^{(i)}$. Equation 4.16 can be also written as

$$\begin{aligned}Q(\theta,\theta^{(i)}) &= \mathrm{E}(\log \mathrm{L}_{X_1,\mathbf{x}_2}(\theta) \mid \mathbf{x}_2, \theta^{(i)})\\ &= \int_{x_1\in A_1} \log \mathrm{L}_{X_1,\mathbf{x}_2}(\theta)\cdot f(x_1;\mathbf{x}_2,\theta^{(i)})\,\mathrm{d}x_1\\ &= \int_{x_1\in A_1} \log f(x_1,\mathbf{x}_2;\theta)\cdot f(x_1;\mathbf{x}_2,\theta^{(i)})\,\mathrm{d}x_1\ . \end{aligned} \tag{4.18}$$

It should be noted that two arguments in the function $Q(\theta, \theta^{(i)})$ have different meanings. The first argument θ presents the parameter (or set of parameters) that should be gained as a result of maximization of the likelihood function. On the other hand, the second argument $\theta^{(i)}$ corresponds to the parameter that is used in i-th iteration for the evaluation of the expectation. This implies that the maximization in Equation 4.17 is one dimensional maximization problem that can be easily solved by θ when $\theta^{(i)}$ is known.

The convergence of EM algorithm is provided by the following theorems.

Theorem 4.1. [8] If $\theta^{(i+1)} = \underset{\theta}{\operatorname{argmax}}\, Q(\theta, \theta^{(i)})$, then

$$\mathrm{L}_{X_1,\mathbf{x}_2}(\theta^{(i+1)}) \geq \mathrm{L}_{X_1,\mathbf{x}_2}(\theta^{(i)})$$

with equality holding if and only if

$$Q(\theta^{(i+1)}, \theta^{(i)}) = Q(\theta^{(i)}, \theta^{(i)}) \ .$$

Theorem 4.2. [8] Suppose that $\theta^{(i)}$ for $i = 0, 1, 2, \ldots$ is an instance of EM algorithm such that

1. the sequence $\mathrm{L}_{X_1,\mathbf{x}_2}(\theta^{(i)})$ is bounded, and
2. $Q(\theta^{(i+1)}, \theta^{(i)}) - Q(\theta^{(i)}, \theta^{(i)}) \geq \alpha \cdot (\theta^{(i+1)} - \theta^{(i)})(\theta^{(i+1)} - \theta^{(i)})^{\mathrm{T}}$ for some scalar $\alpha > 0$ and all i.

Then the sequence $\theta^{(i)}$ converge to some $\theta^\star$ in the closure of Θ.

The first theorem says that in each step of EM algorithm the likelihood function is non-decreased, and the second theorem implies that the limit $\theta^\star$ of the sequence $\{\theta^{(i)}\}_{i\in\mathbb{N}}$ will be a **local maximum** of $\mathrm{L}_{X_1,\mathbf{x}_2}(\theta)$. However, it is possible that $\theta^\star$ is a global maximum, but there is no guarantee. The proofs of the theorems are given in [8] together with the corollaries and other properties of the EM algorithm.

4.3.2 EM for Poisson-LogNormal Aggregated Loss Distribution

Following the notation from the previous section a complete loss data $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$ is assumed to be a n-times i.i.d. random realization of loss event for one event type and one business line in OR matrix. Namely, our attention is mainly concentrated on obtaining the aggregated distribution function for that cell in the matrix. Let recall that in Section 3. the aggregated loss distribution for total loss $L_{i,j}$ in (i, j)

cell is defined by Equation 3.4. This time the indices i and j referee to number of observations $i = 1, 2, \ldots, n_j$ in A_j set where $j = 1, 2$. The realization of total loss random variable L in A set is

$$l = \sum_{j=1}^{2} \sum_{i=1}^{n_j} x_{ji} \ .$$

Regarding the choices of distributions the frequency of X is assumed to follow Poisson distribution with intensity rate $\Delta t \lambda$, $\Delta t = T_2 - T_1$, and the severity of X is assumed to follow LogNormal distribution $f(x; \mu, \sigma)$. For simplicity, parameters μ and σ will be presented by vector parameter θ. Further, the Assumption 1 and Assumption 2 under standard LDA model also hold.

Now, the aggregated loss density function $g(\mathbf{x}; \lambda, \theta)$ for the complete data set is expressed as

$$g(\mathbf{x}; \lambda, \theta) = \frac{(\Delta t \lambda)^{n_1+n_2}}{(n_1+n_2)!} e^{(-\Delta t \lambda)} \binom{n_1+n_2}{n_1} q_1^{n_1} q_2^{n_2} \prod_{j=1}^{2} \prod_{i=1}^{n_j} \frac{f(x_{ji}; \theta)}{q_j} \tag{4.19}$$

where q_j is the probability that realization of random variable X fall into A_j set i.e. $q_j = \mathrm{P}(x_{ji} \in A_j)$ for $j = 1, 2$. Naturally, the sum of probabilities q_1 and q_2 is equal to 1.

The first part of expression in Equation 4.19 comes from Poisson distribution considering the whole set A. The number of observations is $n = n_1 + n_2$ and the intensity parameter is $\Delta t \lambda$. Since we do not know the number n_1 it is supposed that it is distributed by Binomial distribution. This means that n_1 is chosen from n with probability q_1. The last term in the observed equation considers the severity of X. Since the realization of X can fall in one of two sets A_1 or A_2, the distribution function f is adjusted by that corresponding probabilities q_1 or q_2. Consequently, it follows that aggregated loss distribution g depends, not only on frequency parameter λ and severity vector parameter θ, but also on probabilities q_j, $j = 1, 2$. Moreover, since

$$q_j = \mathrm{P}(x_{ji} \in A_j) = \mathrm{P}(x_{ji} \leq u) = F(u; \theta)$$

it follows that g depends, also, on the values of threshold u.

The simplification of Equation 4.19 gives

$$g(\mathbf{x}; \lambda, \theta) = \frac{(\Delta t \lambda)^{n_1+n_2}}{n_1! \ n_2!} e^{(-\Delta t \lambda)} \prod_{j=1}^{2} \prod_{i=1}^{n_j} f(x_{ji}; \theta) \ . \tag{4.20}$$

Generally speaking, the Expectation-Maximization (EM) algorithm is performed in the following way.

1. Setting the Parameters. The algorithm starts with choosing the initial values for unknown parameters λ and θ. Let denote them as $(\lambda^{(0)}, \theta^{(0)})$. Further, having the initial value $\theta^{(0)}$ we can calculate q_j probabilities, $j = 1, 2$ as $q_1^{(0)} = F(u; \theta^{(0)})$ and $q_2^{(0)} = 1 - q_1^{(0)}$. Then, the initial value for the unknown number n_1 is computed from the expression $n_1^{(0)} = q_1^{(0)} \lambda^{(0)} \Delta t$. This follows from equalities $\lambda = \lambda_1 + \lambda_2 = q_1 \lambda + (1 - q_1)\lambda = \frac{n_1 + n_2}{\Delta t}$.

2. E-step. In this step the conditional expectation of log-likelihood function of complete random sample $\mathbf{x}$ given the known sample $\mathbf{x}_2$ and initial parameters $(\lambda^{(0)}, \theta^{(0)})$ is calculated. That is, we need to obtain the following expectation

$$\mathrm{E}(\log \mathrm{L}_{X_1, \mathbf{x}_2}(\lambda, \theta); \mathbf{x}_2, \lambda^{(0)}, \theta^{(0)}) \ . \tag{4.21}$$

In our case the log-likelihood function of aggregated loss density is equal to

$$\begin{aligned} \log \mathrm{L}_{X_1, \mathbf{x}_2}(\lambda, \theta) &= (n_1 + n_2) \log(\Delta t \lambda) - \Delta t \lambda - \\ &- \sum_{j=1}^{n} \log(n_j!) + \sum_{j=1}^{2} \sum_{i=1}^{n_j} \log f(x_{ji}; \theta) \end{aligned} \tag{4.22}$$

and requesting conditional expectation is equal to

$$\begin{aligned} \mathrm{E}(\log \mathrm{L}_{X_1, \mathbf{x}_2}(\lambda, \theta); \mathbf{x}_2, \lambda^{(0)}, \theta^{(0)}) &= (n_1^{(0)} + n_2) \log(\Delta t \lambda) - \Delta t \lambda + \\ &+ n_1^{(0)} \mathrm{E}(\log f(X_1, X_2; \theta); \mathbf{x}_2, \theta^{(0)}) + \\ &+ \sum_{i=1}^{n_2} \log f(x_{2i}; \theta) \ . \end{aligned} \tag{4.23}$$

3. M-step. The next step is maximization of Equation 4.23 with respect to unknown parameters λ and θ, that is

$$(\lambda^{(1)}, \theta^{(1)}) = \underset{\lambda, \theta}{\operatorname{argmax}} \, \mathrm{E}(\log \mathrm{L}_{X_1, \mathbf{x}_2}(\lambda, \theta); \mathbf{x}_2, \lambda^{(0)}, \theta^{(0)}) \ . \tag{4.24}$$

In our case

$$\lambda^{(1)} = \frac{n_1^{(0)} + n_2}{\Delta t} \tag{4.25}$$

$$\begin{aligned} \theta^{(1)} &= \underset{\theta}{\operatorname{argmax}} \Big(n_1^{(0)} \cdot \mathrm{E}\big(\log f(X_1, X_2; \theta); \mathbf{x}_2, \theta^{(0)} \big) + \\ &+ \sum_{i=1}^{n_2} \log f(x_{2i}; \theta) \Big) \end{aligned} \tag{4.26}$$

where the expectation in (4.26) is equal to the following integral

$$\mathrm{E}\big(\log f(X_1, X_2; \theta); \mathbf{x}_2, \theta^{(0)}\big) = \int\limits_{x_1 \in A_1} \log f(x_1, \mathbf{x}_2; \theta) f(x_1; \mathbf{x}_2, \theta^{(0)}) \mathrm{d}x_1 \ . \tag{4.27}$$

4. Convergence. For the new obtained values $\lambda^{(1)}$ and $\theta^{(1)}$ new E-step and M-step are repeated until convergence is reached.

4.4 Statistical Goodness-Of-Fit Tests

Suppose that a data sample $\mathbf{x} = (x_1, \ldots, x_n)$ has been fitted by some parametrical distribution function $F(x; \theta)$ for which the parameter θ is estimated and denoted as $\hat{\theta}$. In order to determine the goodness of that fit one need to perform a statistical test. There is quite a number of Goodness-Of-Fit (GOF) tests that provide a comparatione of the empirical distribution function with the cumulative distribution function used for fitting the data. For example, Kolmogorov-Smirnov test, Anderson-Darling test, Cramer-von Mises test etc. In our empirical study the Kolmogorov-Smirnov GOF test (K-S) was considered.

4.4.1 Kolmogorov-Smirnov GOF test

The testing hypothesis for every GOF test are the following

$$\mathrm{H}_0 : F_{ecdf}(x) \in F(x; \hat{\theta}) \quad \text{vs.} \quad \mathrm{H}_1 : F_{ecdf}(x) \notin F(x; \hat{\theta}), \quad \forall x \in \mathbf{x}$$

where F_{ecdf} is the empirical cumulative distribution function explained earlier by Equation 4.1.

The K-S test reject or accept the H_0 hypothesis according to the absolute difference between the empirical and fitted distribution function. Firstly, $\mathbf{x}$ data set must be put in a vector of order statistics, that is $x_{(1)} < x_{(2)} < \ldots < x_{(n)}$. Then, the KS statistic value is calculated as

$$\mathrm{KS} = \sup_{1 \le i \le n} \mid F_{ecdf}(x_{(i)}) - F(x_{(i)}; \hat{\theta}) \mid \ . \tag{4.28}$$

Consequently, in this GOF test every observation in data sample has the same weight. This means that K-S test does not consider where the maximal difference occurred only how big the difference is.

The rejection of H_0 hypothesis depends mostly on the computed p-value. The p-value is the probability that KS statistic value is greater than ks critical value for the given confidence level α i.e.

$$p\text{-value} = \mathrm{P}(\mathrm{KS} > \mathrm{ks}) \ .$$

Usually, confidence level α goes from 5% to 10% and the corresponding critical values are $1.36/\sqrt{n}$ and $1.22/\sqrt{n}$. So, if p-value is lower than 0.05 or 0.1 then H_0 is rejected.

This K-S GOF test is used for data sets which are considered not-truncated or not-censored. Consequently, in our empirical study this test is applied on the data sets that are fitted by distributions according to the standard approach from Section 4.3.

4.4.2 K-S test for Left-Truncated Data

In order to test the goodness of fit for left-truncated data sets below some threshold u, the standard K-S test need to be modified.

Using the old notation, the left-truncated data set $\mathbf{x}_2 = (x_{21}, x_{22}, \ldots, x_{2n_2}) \in (u, \infty)^{n_2}$ is expanded by the missing data set $\mathbf{x}_1 = (x_{11}, x_{12}, \ldots, x_{1n_1}) \in [0, u]^{n_1}$. Let recall that under the adjusted approach from Section 4.3 the frequency and severity distributional parameters are computed as

$$\hat{\tau}^{adj} = \frac{\hat{\tau}^{mle}}{1 - F(u; \hat{\theta}^{adj})} \tag{4.29}$$

$$\hat{\theta}^{adj} = \underset{\theta}{\operatorname{argmax}} \log \Big(\prod_{k=1}^{n_2} \frac{f(x_{2k}; \hat{\theta})}{1 - F(u; \hat{\theta})} \Big) \tag{4.30}$$

Note that the loss frequency density function $P(n; \hat{\tau}^{mle})$ is not changed. Namely, only the parameter $\hat{\tau}^{mle}$ estimated by Maximum Likelihood (MLE) method is adjusted by (4.29), and thus, the adjusted frequency density function is simply $P(n; \hat{\tau}^{adj})$. As far as loss severity density function is concerned, the modification is done in the following way. The complete (true) density function f is adjusted for the truncated data what results as a new truncated distribution function $f^{adj}(x; \hat{\theta}^{adj})$. As mentioned before, its density function is defined as

$$f^{adj}(x; \theta) = \begin{cases} \dfrac{f(x; \theta)}{1 - F(u; \theta)} & , \quad x \geq u \\[2ex] 0 & , \quad x < u \end{cases} \tag{4.31}$$

and its cumulative distribution as

$$F^{adj}(x;\theta) = \begin{cases} \dfrac{F(x;\theta) - F(u;\theta)}{1 - F(u;\theta)} & , \quad x \geq u \\ \\ 0 & , \quad x < u \end{cases} \tag{4.32}$$

The total number of observations for complete data set $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$ is equal to $n = n_1 + n_2$ and the expected number of data depends on the fitted distribution. If we choose Poisson distribution, the expected number is equal to intensity parameter λ. However, λ for $\mathbf{x}_2$ data set is $\hat{\lambda}_2 = \frac{n_2}{\Delta t}$ and accordingly, for the $\mathbf{x}$ it is $\hat{\lambda} = \frac{n}{\Delta t}$. Namely, since we have two area, λ can be expressed as

$$\begin{aligned} \lambda &= \mathbf{P}(x \leq u) \cdot \lambda + \mathbf{P}(x > u) \cdot \lambda \\ &= F^{adj}(u;\theta) \cdot \lambda + (1 - F^{adj}(u;\theta)) \cdot \lambda \, . \end{aligned}$$

Further, we obtain

$$\begin{aligned} \lambda_1 &= F^{adj}(u;\theta) \cdot \lambda = \frac{n_1}{\Delta t} \\ \lambda_2 &= (1 - F^{adj}(u;\theta)) \cdot \lambda = \frac{n_2}{\Delta t} \end{aligned}$$

From the above equalities it follows that number of missing data and number of observed data is related as

$$n_1 = \frac{n_2 \, F^{adj}(u;\theta)}{(1 - F^{adj}(u;\theta))} \tag{4.33}$$

meaning that total number n is

$$n = \frac{n_2}{1 - F^{adj}(u;\theta)} \, . \tag{4.34}$$

This relation implies that the empirical cumulative distribution $F_{ecdf}(x)$ of complete data set $\mathbf{x}$ can not be based only on the number of observed data n_2. Namely, the empirical cumulative distribution $F^{n_2}_{ecdf}$ of $\mathbf{x}_2$

$$F^{n_2}_{ecdf}(x) = \frac{1}{n_2} \sum_{i=1}^{n_2} \mathrm{I}(x_{2i} \leq x)$$

explains only the truncated data set $\mathbf{x}_2$. Since all missing data from $\mathbf{x}_1$ are by definition lower than threshold u it means that $F_{ecdf}(x)$ can be expressed using

Equations 4.34 and 4.33 respectively as

$$\begin{aligned} F_{ecdf}(x) &= \frac{n_1 + \sum_{i=1}^{n_2} \mathrm{I}(x_{2i} \le x)}{n} \\ &= \frac{\big(1 - F^{adj}(u;\theta)\big)\big(n_1 + \sum_{i=1}^{n_2} \mathrm{I}(x_{2i} \le x)\big)}{n_2} \\ &= F^{adj}(u;\theta) + (1 - F^{adj}(u;\theta)) \cdot \frac{1}{n_2} \sum_{i=1}^{n_2} \mathrm{I}(x_{2i} \le x) \end{aligned}$$

what is further equal to

$$F_{ecdf}(x) = F^{adj}(u;\theta) + (1 - F^{adj}(u;\theta)) \cdot F_{ecdf}^{n_2}(x) \ . \tag{4.35}$$

Now, having expression for adjusted cumulative distribution (4.32) and for complete empirical cumulative distribution (4.35) the testing hypothesis are

$$\mathrm{H}_0 : F_{ecdf}(x) \in F^{adj}(x;\theta) \quad \text{vs.} \quad \mathrm{H}_1 : F_{ecdf}(x) \notin F^{adj}(x;\theta), \qquad \forall x \in \mathbf{x}$$

Accordingly, the KS statistic value is also modified as

$$\mathrm{KS}^{\star} = \sqrt{n} \sup_{1 \le i \le n} \mid F_{ecdf}(x_i) - F^{adj}(x_i;\hat{\theta}) \ . \tag{4.36}$$

The calculation of p-value under these hypothesis must be done in the different way. This is due to the fact that now our empirical cdf is dependant on the parameter θ from F^{adj} distribution. One way of determining the p-value is by Monte-Carlo simulation (see A. Chernobai 2005, [20]). The simulation steps are the following.

1. Determine the confidence level α. In our case it is $\alpha = 0.05$.
2. Calculate the $\mathrm{KS}^{\star}$ for $\mathbf{x}$ data set.
3. Obtain the large number of random samples y_m, $m = 1, \ldots, M$ of dimension n from the $F^{adj}(x;\theta)$ with condition $y_m > u$. The number of random samples should be 1,000 or higher.
4. Fit every random sample y_m with $F^{adj}(y_m;\theta)$ and estimate the corresponding parameter $\hat{\theta}_m^{adj}$.
5. Calculate the values of KS statistics for every random sample y_m and denote them as $\mathrm{KS}_m^{\star}$.

6. The p-value is computed as number of $\mathrm{KS}_m^\star$ exceeding the $\mathrm{KS}^\star$, that is

$$p\text{-value} = \frac{\sum_{m=1}^{M} \mathrm{I}(\mathrm{KS}_m^\star > \mathrm{KS}^\star)}{M} \; .$$

7. If p-value is lower than α then H_0 hypothesis is rejected.

In paper [20], the other GOF tests which can be used for testing the left-truncated data samples are introduced. The basic idea is the same as explained above.

4.5 Extreme Value Theory

The most difficult problem in the OR analysis is the question of modelling extremely rare and extremely high operational losses, that is, how to treat large and rare losses which come from low-frequency high-impact events?

Over the recent years, researchers have started to explore whether Extreme Value Theory (EVT), as a set of statistical techniques, can be used in measuring OR exposure. It appears that for the tail of the data EVT is an useful method. It can handle the modelling of large losses and estimation of high quantiles of a loss distribution. Particulary, the extreme (limit) distributions from the EVT theory are considered.

EVT is a scientific approach and has solid foundations in the mathematical theory of the behavior of extremes, as the name says. It was set up by Fisher and Tippet in 1928, [18]. The theory is focused only on analysis of the tail area and, hence, reduce the influence of the small/meduim-sized losses. Hence, EVT does not need all loss data, it requires only large loss data. Moreover, EVT is applied in various fields such as structural engineering, aerospace, ocean and hydraulic engineering, material strength, studies of pollution, meteorology, highway traffic, and, more recently, in the financial and insurance fields. Since the original underlying distribution of all data is generally unknown, what is good about EVT is that it does not require some particular assumptions on the nature of the underlying distribution.

There are two related approaches in which EVT is applied. The first one is the Block Maxima and the second one is the Peaks Over Threshold (POT).

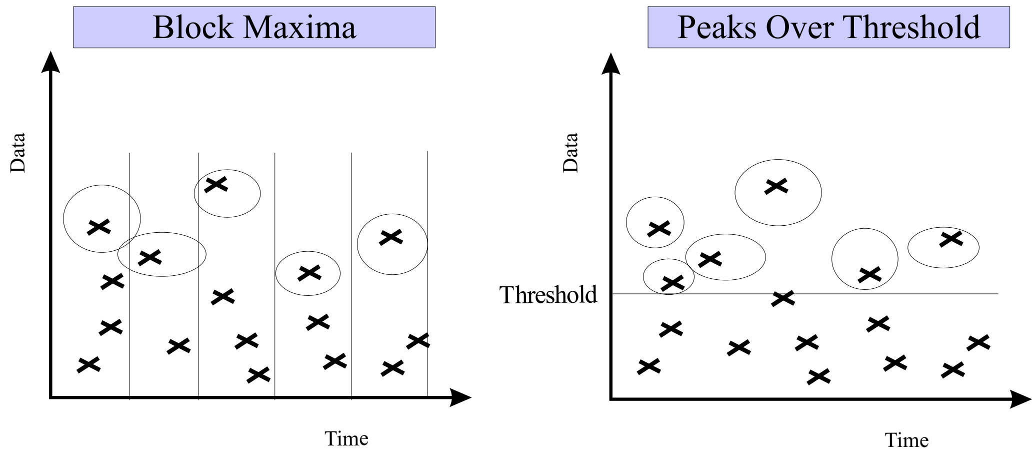

Figure 4.6: The EVT Methods

4.5.1 Block Maxima

Block Maxima is a basic approach considers the maxima (or minimum) values that random variable (e.g. operational loss data) takes in successive periods, for example months or years. To be more precise, EVT deals with the limiting distribution of sample extremes. See Figure 4.6.

Let denote a sequence of independent identically distributed (i.i.d.) random variables by $X = (X_1, \ldots, X_n)$ with a cumulative distribution function $F_X(x)$. As usual $F_X(x)$ is unknown distribution function which describe the behavior of OR data for a considered business line or event type. Further, $M_n = \max\{X_1, \ldots, X_n\}$ is the sample maximum also called extreme event, block maxima or per-period maxima.

The "three-types theorem" by Fisher and Tippet, [18] states that there are only three types of distributions which can arise as limiting distributions of extreme values in random samples. These three types of extremes distributions are: **Weibull**, **Gumbel** and **Frechet** type. This implies that the asymptotic distribution of the maxima M_n always belongs to one of these three distributions, regardless of the original one i.e. the distribution F_X.

The Weibull, Gumbel and Frechet distributions can be represented as the **Generalised Extreme Value distribution**(GEV)

$$GEV_{\xi,\mu,\sigma}(x) = \begin{cases} \exp(-(1+\xi\frac{x-\mu}{\sigma})^{-\frac{1}{\xi}}) & \text{if } \xi \neq 0 \\ \\ \exp(-\exp(-\frac{x-\mu}{\sigma})) & \text{if } \xi = 0 \end{cases} \tag{4.37}$$

where $1 + \xi x > 0$. Following, from GEV for different values of ξ three type of distribution can be obtained

- Weibull distribution $\Psi_\alpha : \xi = -\alpha^{-1} < 0$
- Gumbel distribution $\Lambda : \xi = 0$
- Frechet distribution $\Phi_\alpha : \xi = \alpha^{-1} > 0$

The parameters μ and σ correspond to location and scale while ξ the third parameter (the shape index) indicates the thickness of the tail of the distribution. The smaller the shape index ξ is, the thicker the tail is.

Obviously, distributions can be classified according to their tail thickness. There are light-tailed distributions, medium-tailed distributions and heavy-tailed distributions. However, there is no common agreed-upon definition but one of the definition is based on a distribution's maximal moment. Namely, if the maximal moments $\sup_r \{E(X^r) < \infty\}$ are finite of all orders than the distribution is considered light-tailed, and heavy-tailed otherwise. So concretely,

- *Light-tailed distribution* ($\xi < 0$) has all finite moments, is characterised by finite right endpoint in tails and converge to the Weibull curve (e.g. Weibull, Uniform, Beta);
- *Medium-tailed distribution* ($\xi = 0$) is also with all finite moments, but the cumulative distribution functions decline exponentially in the tails like the Gumbel curve (e.g. Normal, Gamma, LogNormal, Exponential);
- *Heavy-tailed distribution* ($\xi > 0$) has the cumulative distribution functions which decline like a power function $x^{-\frac{1}{\xi}}$ in the tails like Frechet curve (e.g. Pareto, LogGamma, GPD, Burr, LogLogistic, T-Student).

The purpose of tail estimation approach is to estimate the random values X outside the range of existing data. To do this, researchers have taken into consideration both extreme events, and exceedances over the specified threshold.

4.5.2 Peaks Over Threshold

Peaks over threshold (POT), the second approach to EVT, considers only data which are bigger than a given (chosen) high threshold. See Figure 4.6.

In modelling the severity of operational losses POT method use a two parameter distribution **Generalised Pareto Distribution (GPD)**, with cumulative function expressed as:

$$GPD_{\xi,\sigma}(x) = \begin{cases} 1-(1+\xi\dfrac{x}{\sigma})^{-\frac{1}{\xi}} & , \quad \xi \neq 0 \\ \\ 1-\exp(-\dfrac{x}{\sigma}) & , \quad \xi = 0 \end{cases} \tag{4.38}$$

where

$$x \in \begin{cases} [0,\infty] & , \quad \xi \geq 0 \\ (0,-\frac{\sigma}{\xi}] & , \quad \xi < 0 \end{cases}$$

The parameter σ is the scale parameter while ξ is the shape index.

The GPD distributions can be extended by adding a location parameter μ. In that case the GPD is defined as:

$$GPD_{\xi,\mu,\sigma}(x) = \begin{cases} 1-(1+\xi\dfrac{x-\mu}{\sigma})^{-\frac{1}{\xi}} & , \quad \xi \neq 0 \\ \\ 1-\exp(-\dfrac{x-\mu}{\sigma}) & , \quad \xi = 0 \end{cases} \tag{4.39}$$

The interpretation of shape index ξ in the GPD is the same as in the GEV. Namely, all relevant information on the tail of the original (unknown) distribution F is embedded in this parameter, since POT method considers only the data that exceed the threshold and those are the data that constitute the tails of distribution. So concretely, the maxima of samples of events from GPD are GEV distributed with shape parameter equal to the shape parameter of the GPD. Furthermore, there is a simple relationship between the standard GPD and GEV:

$$GPD(x) = 1 + \log GEV(x) \quad \text{if} \quad \log GEV(x) > -1 \ .$$

Same as for GEV, the GPD gives the different distributions for different values of ξ:

- Pareto "Type II" distribution ($\xi < 0$),
- Exponential distribution ($\xi = 0$) and

- Pareto distribution ($\xi > 0$).

It is evident that this distributions together with the shape index ξ coincide with the given classification of distributions based on their tail-thickness.

Consider again the a sequence of i.i.d. random variables $X = (X_1, \ldots, X_n)$ with right-endpoint x_F and with an underlying distribution $F_X(x)$. What draws our attention now is the distribution of excesses over the high threshold u. Y is the excess random variable defined as

$$Y := X - u,$$

where u is a given threshold and the $F_u(y)$ is its excess distribution at the threshold u.

The excess distribution can be viewed and expressed through a conditional distribution function, that is:

$$F_u(y) = P(X - u \leq y; X > u) = \frac{F_X(x) - F_X(u)}{1 - F_X(u)} \ , \tag{4.40}$$

for $y = x - u > 0$.

Obviously, the (4.40) represents the probability of the event that the loss X exceed the u by at most an amount y , if it is given that a loss X exceeds the threshold u.

According to the theory of Balkema-De Haan (1974) and Pickands (1975), for a large class of underlying distributions F, the excess distribution $F_u(y)$ converges asymptotically to a GPD as the threshold tends to the right-endpoint x_F, that is:

$$\lim_{u \to x_F} \sup \mid F_u(y) - GPD_{\xi,\beta}(y) \mid = 0 \ . \tag{4.41}$$

Although, the conditions under which this it true are quite big.

In this case, GPD with two parameters (ξ, β) and the argument y is called "excess GPD" with the following form

$$GPD_{\xi,\beta}(y) = \begin{cases} 1 - (1 + \xi \frac{y}{\beta})^{-\frac{1}{\xi}} & , \quad \xi \neq 0 \\ \\ 1 - \exp(-\frac{y}{\beta}) & , \quad \xi = 0 \end{cases} \tag{4.42}$$

where $y = x - u$ is excess, ξ shape index, β scale, and

$$\begin{array}{lll} y \in [0, x_F - u] & \text{if} & \xi \geq 0, \\ y \in [0, -\beta/\xi] & \text{if} & \xi < 0. \end{array}$$

The limit condition (4.41) holds even if the exceedances x take place of the excesses y. Basically, with this change of the argument, the $F_u(y)$ and $GPD_{\xi,\beta}(y)$ transform respectively to $F_u(x)$ and $GPD_{\xi,u,\beta}(x)$. Therefore, when the threshold tends to the right endpoint x_F, the exceedance distribution $F_u(x)$ converges asymptotically to a GPD with the same shape ξ, scale β and location μ equal to the threshold u ($\mu = u$ and $x > u$). To stress the difference, $GPD_{\xi,u,\beta}(x)$ will be called the "exceedance GPD", since it deals with the exceedances x at u.

Stability of GPD. One of the most important properties of the GPD is its stability under an increase of the threshold.

To show that, let isolate $F_X(x)$ from (4.40)

$$F_X(x) = \Big[1 - F_X(u)\Big] F_u(y) + F_X(u) \ . \tag{4.43}$$

According to limit condition (4.41) and above conclusion from it, both the excess distribution $F_u(y)$ and the exceedance distribution $F_u(x)$ can be well approximated by suitable GPDs. By using the "exceedance GPD", one obtains

$$F_X(x) \approx \Big[1 - F_X(u)\Big] GPD_{\xi,u,\beta}(x) + F_X(u) \ . \tag{4.44}$$

Now, substituting the $GPD_{\xi,u,\beta}(x)$ expression in (4.44) one gets

$$F_X(x) \approx [1 - F_X(u)] \left[1 - (1 + \xi \frac{x-u}{\beta})^{-\frac{1}{\xi}}\right] + F_X(u) \ . \tag{4.45}$$

It is evident, that the only "unknown" element for identification of $F_X(x)$ is $F_X(u)$; that is the value of the (unknown) distribution function in correspondence with the threshold u. One of the possible empirical estimator for $F_X(x)$ computed at level u can be empirical cdf

$$F_n(u) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{I}(x_i \le u) = \frac{n - n_u}{n} \tag{4.46}$$

where n is the total number of observation and n_u the number of observation above the threshold u.

However, the problem of choosing the right threshold u is quite important, since the empirical estimation of $F_X(u)$ requires sufficient number of observations which exceed the given threshold u. Obviously, if there is not enough observations, the estimations could not be reliable.

Further, $F_X(x)$ can be completely expressed by the number of observation (total and over the threshold) and by the parameters of the $GPD_{\xi,u,\beta}$ as

$$F_X(x) \approx \frac{n_u}{n}\left[1-(1+\xi\frac{x-u}{\beta})^{-\frac{1}{\xi}}\right]+(1-\frac{n_u}{n}) \ .$$

Simplified,

$$F_X(x) \approx 1-\frac{n_u}{n}\left[1+(1+\xi\frac{x-u}{\beta})^{-\frac{1}{\xi}}\right] \ . \tag{4.47}$$

This obtained measure (4.47) is called the **"tail estimator"** of $F_X(x)$, as it is valid only for $x > u$. It is possible to demonstrate that the "tail estimator" is also GPD distributed; it is the semiparametric form of the GPD ($GPD_{\xi,\mu,\sigma}$) which refer to all the original data, with the shape index ξ, location μ and the scale σ. Since the $GPD_{\xi,\mu,\sigma}$ is fitted to all data in the tail area, it will be considered as "full GPD". Namely, this semiparametric form of GPD provides information on the frequency of the exceedances (excesses) over the threshold through the $F_n(u)$ which uses the total number and above threshold number of observations, unlike the exceedance GPD ($GPD_{\xi,u,\beta}$) and excess GPD ($GPD_{\xi,\beta}$).

Semiparametric estimates for the full GPD parameters can be derived from those of the exceedance GPD

$$\sigma = \beta\left(\frac{n_u}{n}\right)^{\xi} , \tag{4.48}$$

$$\mu = u - \frac{\beta}{\xi}\left[1-\left(\frac{n_u}{n}\right)^{\xi}\right] . \tag{4.49}$$

Further, the scale parameter β can be expressed as

$$\beta = \sigma + \xi(u-\mu) \ . \tag{4.50}$$

Apparently, it should be noted that, while the scale parameter β of the exceedance GPD depends on where the threshold u is located, the shape index ξ, the location μ and scale σ of the full GPD are *independent of the threshold.* Thus, a nice practical method to check the robustness of the model for some specific data is to evaluate the degree of stability of the parameters (ξ, μ, σ) over a variety of thresholds.

In practice, we should estimate for each chosen threshold u the parameters of exceedance GPD (ξ and β) and after that the corresponding values of the full GPD parameters (μ and σ). Now, having all these parameters one should investigate the approximate equality of ξ, μ and σ for increasing thresholds.

By applying the GPD stability property, it is possible to move easily from the excess data ($y = x - u$) to the tail of the original data ($x > u$) and hence, from the excess distribution $F_u(y)$ to the underlying (unknown) distribution $F_X(x)$.[2]

An immediate consequence of the GPD stability is that if the exceedances of a threshold u follow a $GPD_{\xi,u,\beta}$ the exceedances over a higher threshold v ($v > u$) are $GPD_{\xi,v,\beta+\xi(v-u)}$ distributed. That is, they are also GPD distributed with the same shape index ξ, the location equal to v (the new threshold) and the scale equal to $\beta + \xi(v - u)$ (following from (4.50)).

4.6 Simulation of Aggregated Loss Distribution

It was mentioned before that, in most situations, the analytical representation of aggregated loss distribution G is difficult to get. Thus, the approximation of G is done by means of simulation techniques. Monte Carlo simulation is the most common one, but, also, some others can be found in the literature. Frachot(2001) uses Panjer's recursive approach and inverse of characteristic function introduced by Heckman and Mayers (1983).

In this paper, for the empirical study, the Monte Carlo simulation is adopted, and hence, the statistical background of the method is presented together with applied algorithm.

Monte Carlo Simulation - Theoretical Background. Let assume again that we are given a sequence of i.i.d. random variables $X = (X_1, X_2, \ldots, X_n)$ with density function f. What we want to calculate is the expectation of $g(X)$, where g is some n-dimensional function. When this n-multiple integral

$$E(g(X)) = \int\int \ldots \int_n g(X)f(X)dX$$

can not be computed neither analytically nor numerically, the simulation is the only way for getting the approximation of $E(g(X))$. The simulation begin with random draw from the density function f in order to generate first random vector $\mathbf{x}^{(1)} = (x_1^{(1)}, x_2^{(1)}, \ldots, x_n^{(1)})$. Then, the next step is to calculate $g(\mathbf{x}^{(1)})$. These two steps of generating and computing $g(\mathbf{x}^{(i)})$ are

[2] It is also possible to move from semiparametric full GPD to its completely parametric form, but in this case it is necessary to know all the information on the original data (the amounts of the data under and over the threshold).

done independently for N times. If we denote $Y^{(i)} := g(X^{(i)})$, $i = 1, \ldots, N$ we will get independent and identically distributed random variable Y. Following the strong low of large numbers it holds that

$$\lim_{N\to\infty} \frac{Y^{(1)} + \cdots + Y^{(N)}}{N} = E(Y^{(i)}) = E(g(X)) \ .$$

It follows that the estimate of $E(g(X))$ is the average of generated Y random variables meaning that in OR modelling expected loss can be computed as a mean value of the aggregated distribution.

Monte Carlo Simulation - Algorithm for OR. After we have chosen the frequency and severity distribution functions $P_{i,j}(n)$, $F_{i,j}(x)$ for the (i,j) cell we can apply the Monte Carlo simulation for approximation of the corresponding aggregated loss distribution $G_{i,j}(x)$.

1. The number of loss events in k-th year is drawn at random from the frequency distribution $P_{i,j}$, and is denoted with n_k.
2. From the severity distribution $F_{i,j}(x)$ n_k individual losses are drawn. Let denote them by $(x_1, x_2, ..., x_{n_k})$.
3. The aggregate loss for k-th year is $l^{(k)} = \sum_{i=1}^{n_k} x_i$.
4. Repeat first three steps for N simulated years in order to get sample of aggregated losses $l = (l^{(1)}, l^{(2)}, \ldots, l^{(N)})$.

The result of the above described algorithm is a generated independent and identically distributed random sample of annual aggregated losses.

The number of simulations N should be a very large number, and in the literature it varies from 5,000 to 1 million. In Moscadelli, [19] it is stated that David Lawrence (Citigroup) has showed that for calculation of 99^{th} percentile of aggregated losses 1 million data points are required. In this paper for the empirical study N is also, set to be 1 million.

4.7 Correlation of OR Losses and Capital Charge

Let recall that the standard formula for calculation of total capital charge under LDA is given, as a sum of all capital charges across the OR matrix i.e.

$$\begin{aligned} \mathrm{CC}_\alpha &= \sum_i \sum_j \mathrm{CC}_{i,j;\alpha} \qquad (4.51) \\ &= \sum_i \sum_j \mathrm{VaR}^{i,j}_{\alpha,\Delta t} \ . \end{aligned}$$

Namely, in this section we will investigate the issue of correlation between loss events.

Perfect Correlation. Basically, the formula (4.51) holds only if we assume a perfectly correlation among aggregated losses $L_{i,j}$. This means that losses from different BL-ET combinations would occurred simultaneously in same time during the holding period Δt. Furthermore, it implies that all losses are driven from the same (one) randomness, not from the possible 56 different ones (7 BL $\times$ 8 ET). Of course, this situation is hardly realistic. Thus, the figure of capital charge under this assumption can be overestimated and much higher then the capital charge under Standardized Approach (SA) or Basic Indicator Approach (BIA). This is in contrast to the basic idea of the usage of Advanced Measurement Approach (AMA) models, meaning that AMA capital charge estimate should be much less than computed capital charge under BIA or SA. Moreover, banks are only willing to use the internal models on condition to reduce theirs OR capital charges. Consequently, the problem of correlation of OR loss events is very important and the proposed formula should be taken with caution.

Correlation Effect. The second case deals with the correlated losses. The standard Assumption 1 and Assumption 2 from LDA say that frequency $N_{i,j}$ and severity $X_{i,j}$ are independent random variables and that $X_{i,j}$ is i.i.d. within (i, j) cell. Apparently, it implies that aggregated losses $L_{i,j}$ can have possible 56 source of randomness, and that their correlation is not explained by assumptions. To be more precise, a correlation effect which we investigate is the correlation between different cells, not the one within single cell. The correlation can be examined on the level of severity, frequency and aggregated losses.

Firstly, the independence of frequency of losses for different BL or ET, can be viewed through the behaviour of annual number of losses during the past. Secondly, the dependence of the mean value of the severity random variables can provide information on severity correlation. Therefore, considering the aggregated loss dependence it is naturally driven by frequency and/or severity. For the empirical estimation of correlation degree a sufficient number of historical data is needed. In [12] these issues are investigated. It states that the correlation degree of aggregated losses between different cells is less than 5% under the standard LDA assumptions. Additionally, the non-standard LDA model is introduced where dependance between frequency and severity within one cell is allowed. Even though the correlation degree was not higher than 10%. Consequently, the following formula can be adopted

$$\mathrm{CC}_\alpha = \sum_m \mathrm{EL}_m + \sqrt{\sum_{\substack{m,n \\ m \neq n}} k_{m,n}(\mathrm{CC}_m - \mathrm{EL}_m)(\mathrm{CC}_n - \mathrm{EL}_n)} \tag{4.52}$$

for $m, n = 1, 2, \ldots 56$ cells, where $k_{m,n}$ is a correlation degree set to constant in the range from 5% to 10%.

Chapter 5

Empirical Study

In Section 3 and Section 4 the methodology for implementing the LDA model was explained. In this section the exact way of applying LDA model on a given OR loss data is presented together with results, graphics and conclusions.

The operational loss data used in our empirical study are internal data obtained from one serbian middle size bank. They are collected in time period of three years, from 2003 to 2006.

As mentioned before, OR loss data should be classified according to the loss event types (ET) and business lines (BL) and put in Basel (ET $\times$ BL) OR matrix. In our case only three ET-s and four BL-s have observations as shown in Table 5.1.

	numb. of data	% of total
Event Types		
ET2: External Fraud	15	18.29 %
ET4: Clients, Products & B.Practices	46	56.10 %
ET7: Delivery & Process Management	21	25.61 %
Business Lines		
BL2: Trading & Sales	12	14.63 %
BL3: Retail Banking	31	37.80 %
BL4: Commercial Banking	27	32.93 %
BL5: Payment & Settlement	12	14.63 %
TOTAL	82	100 %

Table 5.1: Number of Observations in ET and BL

The total number of observations is 82 what makes the analysis and modelling quite specific. Namely, from statistical point of view it is hard to provide a good fit of data and accurate parameters estimate in the case where there is less than 100 data points. However in practice, banks need to provide at least some information on theirs risk exposures and they can not wait for the required number of loss events. Obviously, this small number of observations is a limiting factor for analysis but in the same time a realistic situation which most risk managers face, at least in local serbian environment.

The standard LDA approach starts with the modelling of every cell, that is for one ET and one BL, and latter it considers the whole OR matrix. It is evident that in our case small number of data (less than 10 per cell) and missing data for some ET and BL require the adjustment of LDA model. In light of that we have decided to perform two separate analysis, one for ET and one for BL data sets. Following the idea, under the term **data set** we will consider those loss events which come from one business line or one event type depending what analysis is performed.

5.1 Descriptive Analysis

The descriptive statistics provide the basic information about severity of OR loss data in the terms of the mean value, the standard deviation, the skewness, the kurtosis etc.

	min	max	mean	st.dev	skewness	kurtosis
Event Types						
ET2	1,000	1,133,537	129,959	310,588	2.53	8.45
ET4	69	27,291,441	1,178,079	4,125,876	5.72	36.49
ET7	200	2,990,109	255,871	681,057	3.39	13.80
Business Lines						
BL2	69	2,990,109	424,972	876,619	2.35	7.35
BL3	200	1,133,537	74,005	220,693	3.94	18.63
BL4	2,924	27,291,441	1,995,859	5,270,771	4.31	21.20
BL5	1,000	102,765	19,357	31,217	1.87	5.34

Table 5.2: Descriptive Statistics of ET and BL

From Table 5.2 it can be seen that the standard deviation values are much higher than the mean values meaning that there is quite a large number of outliers. Further, from the positive skewness values one can conclude that all data sets are left-skewed meaning that the probability mass is concentrated on the left side of the mean. Moreover, the kurtosis values are greater than 3. It implies that data sets have high level of tail-heaviness.

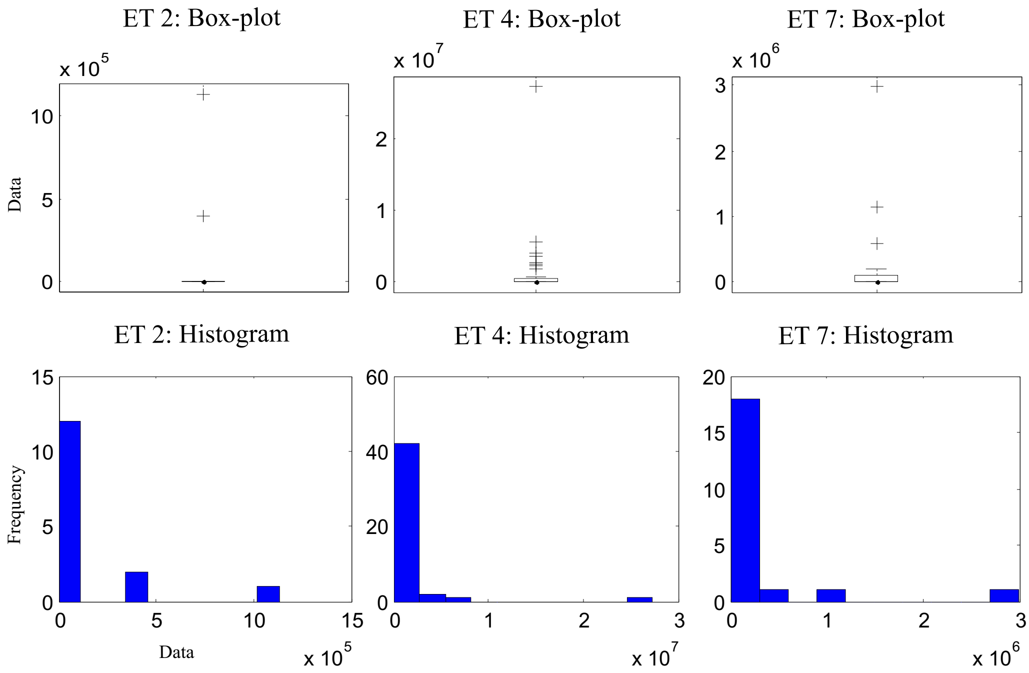

Figure 5.1: ET2, ET4 and ET7: Box-plot, Histogram

Naturally, this findings are in line with box-plots and histograms of data sets. The graphics for ET are presented in Figure 5.1 and for BL in Figures 5.2 and 5.3.

In order to obtain an empirical density function of the data set we have used the Epanechnikov kernel smoothing technique explained in Section 4. These density functions are shown in Figure 5.4 and Figure 5.5. As expected, they suggest the left-skewness property of the data.

Since the tail-heaviness is not clearly viewed from the kernel density function we have also obtained the empirical cumulative density function (cdf). The empirical cdf for the ET data sets are presented in Figure 5.6, and for the BL data sets in Figure 5.7. Apparently, the high kurtosis values imply the heavier tailed cumulative density functions.

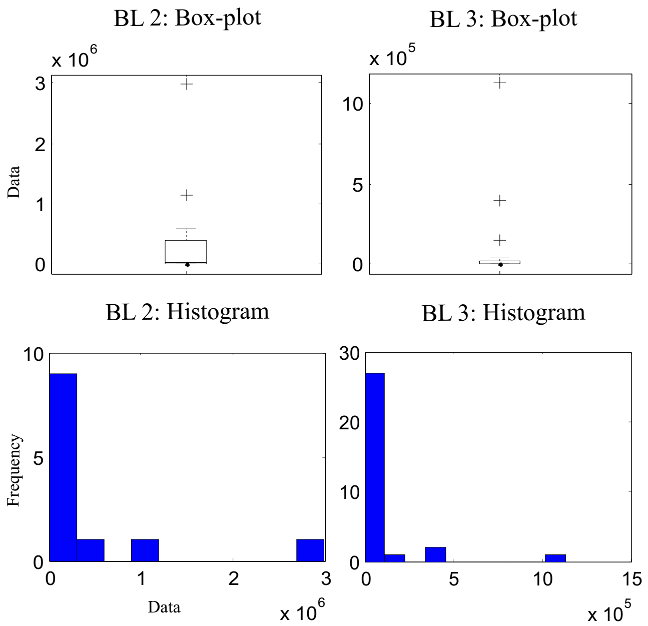

Figure 5.2: BL2 and BL3: Box-plot, Histogram

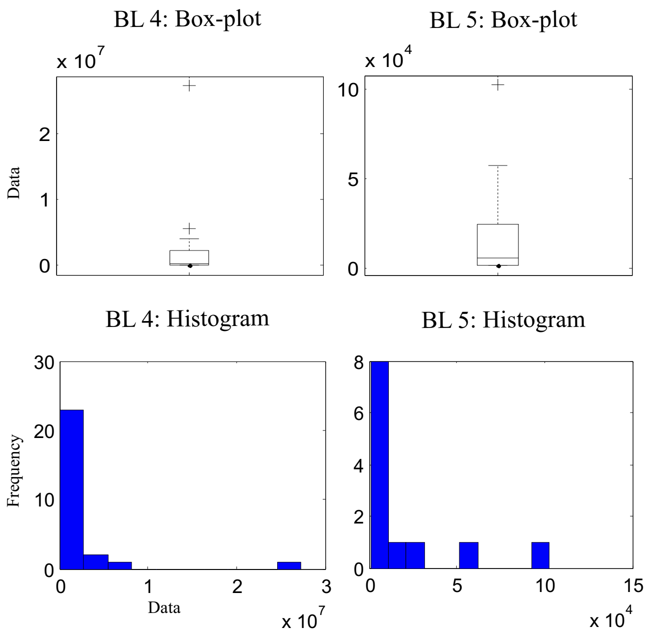

Figure 5.3: BL4 and BL5: Box-plot, Histogram

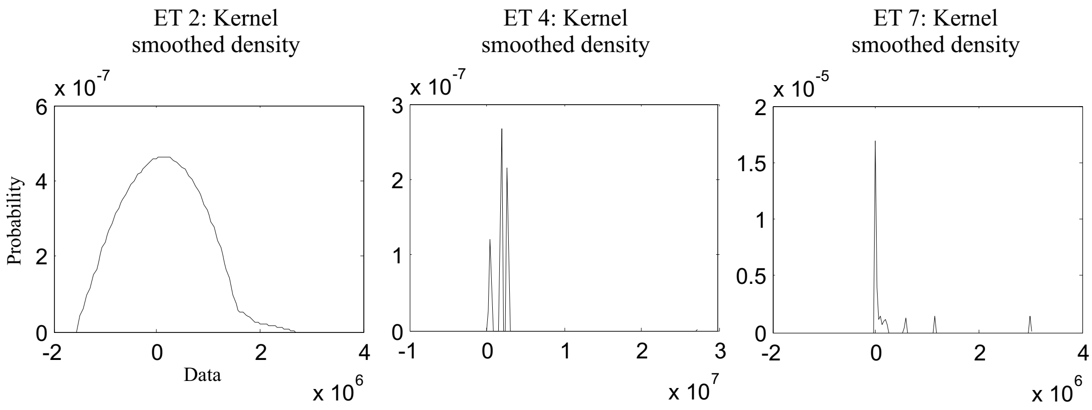

Figure 5.4: ET2, ET4 and ET7: Kernel Smoothed pdf

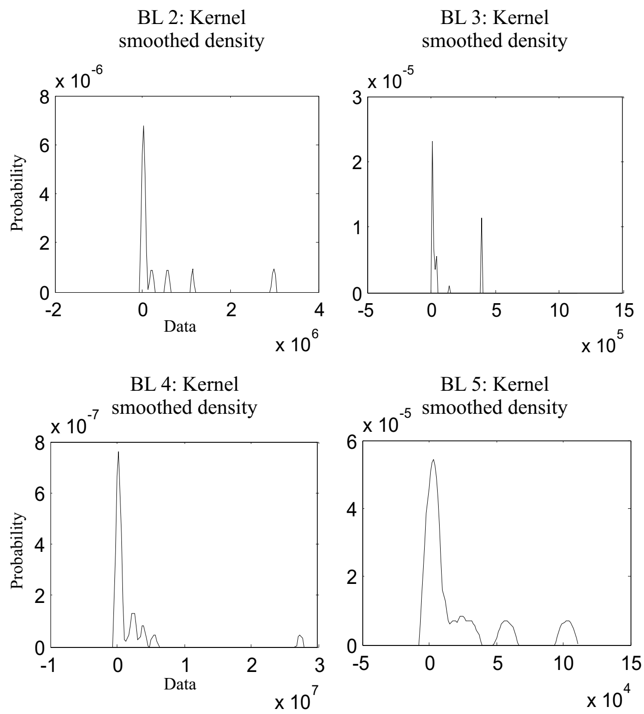

Figure 5.5: BL2, BL3, BL4 and BL5: Kernel Smoothed pdf

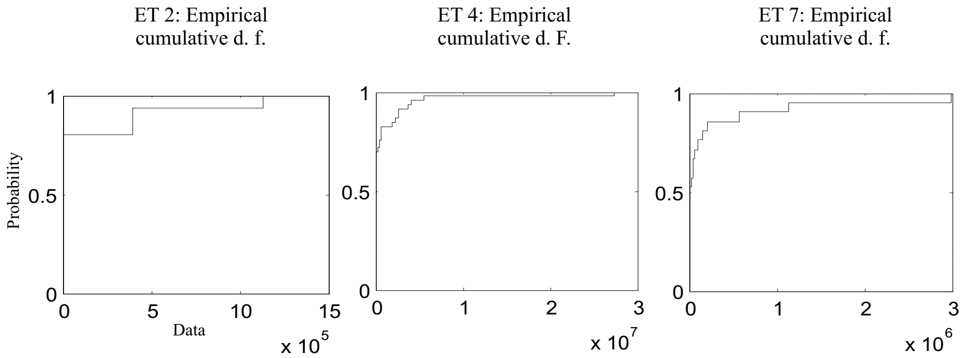

Figure 5.6: ET2, ET4 and ET7: Empirical cdf

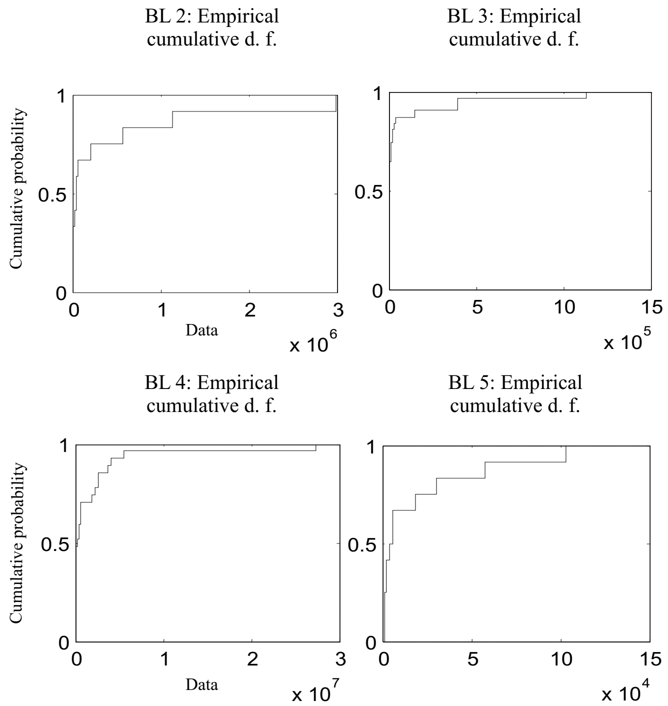

Figure 5.7: BL2, BL3, BL4 and BL5: Empirical cdf

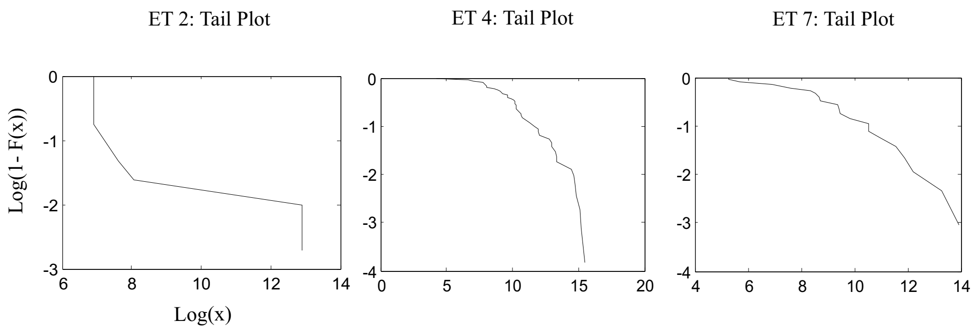

Figure 5.8: ET2, ET4 and ET7: Tail Plot

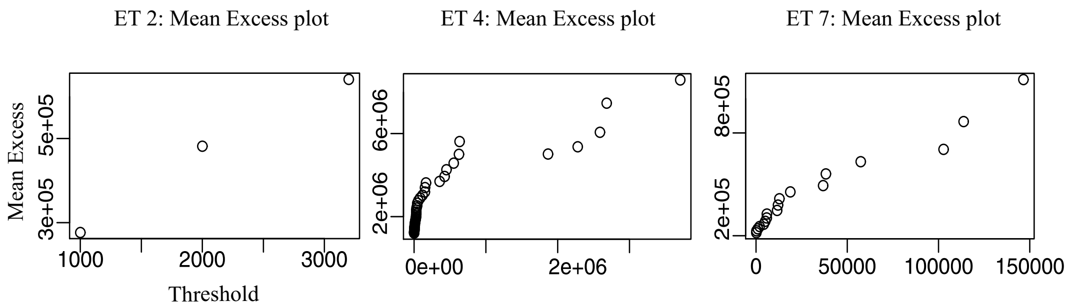

Figure 5.9: ET2, ET4 and ET7: Mean Excess Plot

The last part of the descriptive statistics considers the tail of data sets. It is clear that skewness, kurtosis and above graphics suggest that data sets are more probably driven from heavier tailed and left-skewed distributions. For more information on tail behavior the tail plot and the mean excess plot can be used.

In Figure 5.8 and Figure 5.10 the tail plots for event types and for business lines are shown, respectively. Since all plots are above the reference line of the slope -1 once again it follows that all data sets are driven from heavier tailed distributions.

The mean excess plots for the given data sets are presented in Figure 5.9 and Figure 5.11, and they provide the same conclusion. That is, all plots have a positive slope meaning that data sets belong to heavier tailed distributions.

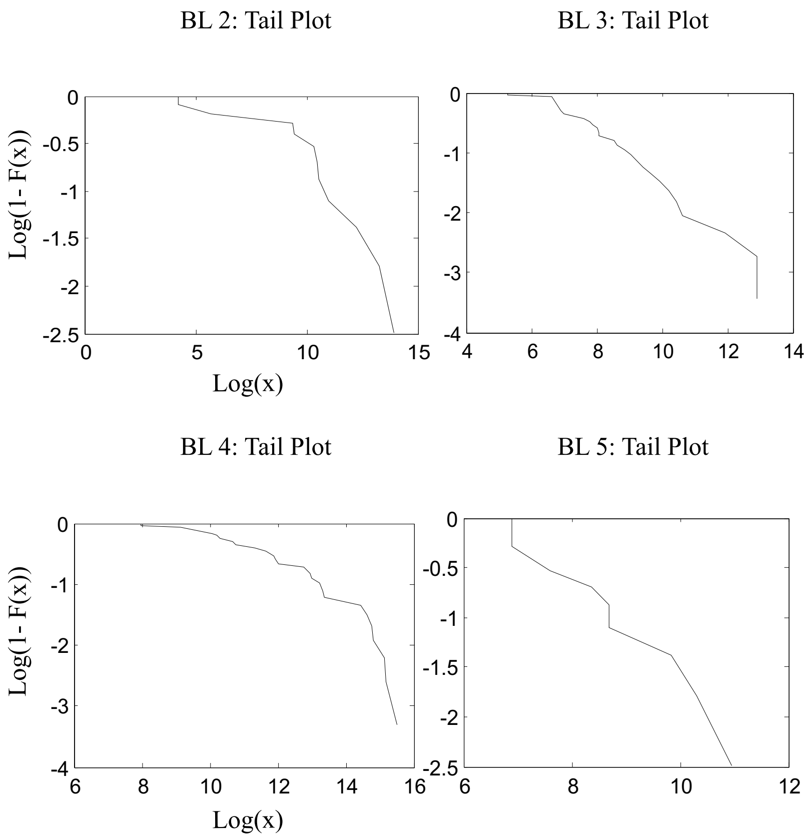

Figure 5.10: BL2, BL3, BL4 and BL5: Tail Plot

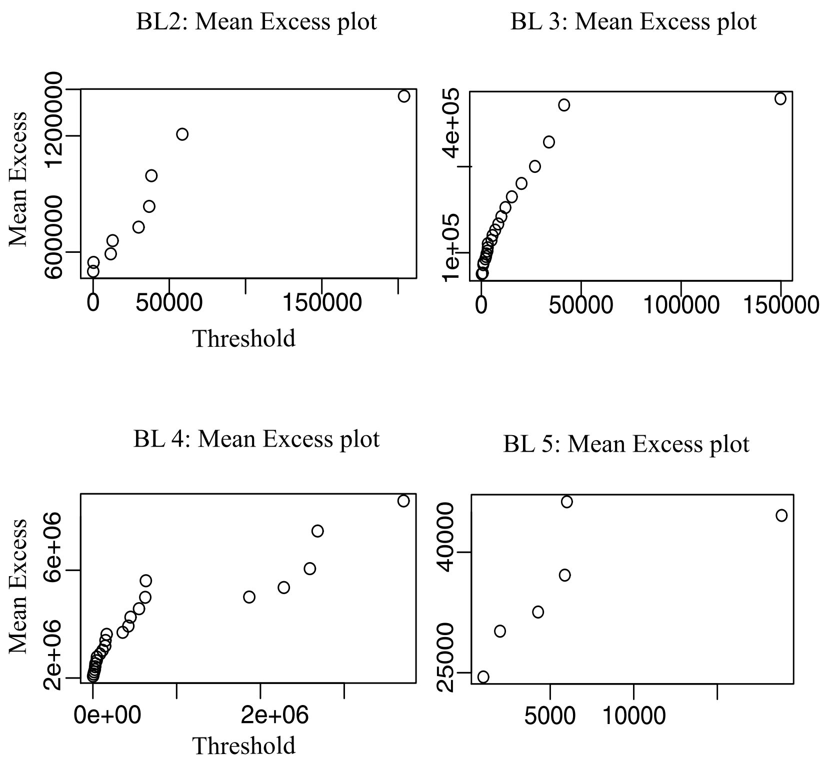

Figure 5.11: BL2, BL3, BL4 and BL5: Mean Excess Plot

Apparently, all graphics and measures are in line and suggest a heavier tailed property of severity of data sets. This can cause problems in finding the most suitable distribution of data since there is a small number of distributions that could fit well and explain the tail of this kind of data. The next section deals with these issues.

5.2 Fitting the Data

Before we start to fit distributions to data sets it is important to note the following characteristic of given OR loss data. Namely, examining the data we have noticed that the minimum recorded loss varies within different data sets. Naturally, one explanation is that loss events occurred in one business line/event type can be more sever than in some other. However, it can also be due to the fact that the bank set some threshold below which the data are not fully recorded. Moreover, this threshold could be different for different data sets. In fact, in every bank the process of collecting and recording the OR loss events is considerably different. The threshold above which the loss events are recorded is a subject of risk managers' decisions. Therefore, we wanted to investigate the influence of different threshold for our data sets. In light of that we have constructed two sets.

A-set is the first considered set that includes all loss events that the bank has provided. The minimum loss event is 69 and it is set to be the threshold for all data sets.

B-set is the second observed set where the threshold for every BL and ET is set to be 3,000. This means that all loss events which are less than 3,000 have been excluded from the data sets.

Now, consequently, in our analysis every BL or ET data set have A-set with small threshold and B-set with high threshold. Table 5.1 presents the numbers of loss events for A-sets while Table 5.3 for B-sets. Let recall that A-set has in total 82 observations while B-set has 61. It means that 25.61 % of data were excluded.

	numb. of data	% of total
B-set		
ET2	4	6.56 %
ET4	40	65.57 %
ET7	17	27.87 %
BL2	10	16.39 %
BL3	18	29.51 %
BL4	26	42.62 %
BL5	7	11.48 %
TOTAL	61	100 %

Table 5.3: Number of Observations for B-set in ET and BL

The first aim in LDA model is to find the most appropriate distribution functions for the frequency and severity of the data set. In Section 4 the parametrical distributions used in the fitting process are listed. For fitting a distribution to the frequency of data set we have used two distributions: Poisson and Negative Binomial. For fitting a distribution to the severity of data set we have used four distributions: Weibull (light-tailed distribution), LogNormal, LogLogistic (medium-tailed) and Pareto (heavy-tailed distribution).

Computation of distributions' parameters is performed in two ways, standard and adjusted approach, as explained in Section 4.

5.2.1 Standard Approach

Under the standard approach the method of Maximum Likelihood Estimation (MLE) is performed. In the following tables the parameters estimates are reported together with results from Kolmogorov-Smirnov goodness of fit test(K-S) (i.e. the values of K-S test and the corresponding p-values). A level of significance or confidence level is 95%.

Event Types - Frequency fitting

Starting from ET data sets, in Table 5.4 and Table 5.5 the fitting results for Poisson and Negative Binomial distribution function are presented, respectively.

Poisson distribution function

	Parameter estimate	K-S test results	
	λ	K-S value	p-value
A-set			
ET2	5	0.282	0.931
ET4	15.333	0.425	0.519
ET7	7	0.584	0.162
B-set			
ET2	1.333	0.286	0.923
ET4	13.333	0.328	0.824
ET7	5.666	0.483	0.358

Table 5.4: ET Frequency - Results of Poisson Fitting

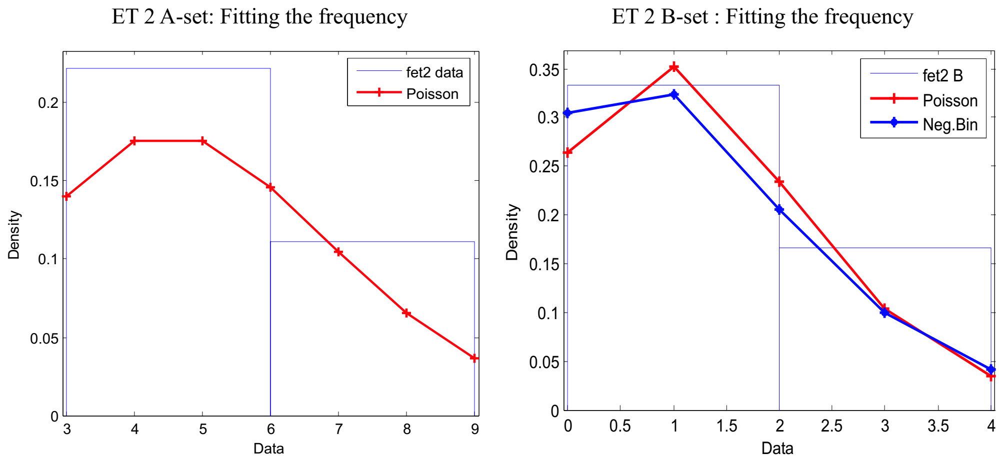

Figure 5.12: ET2: Frequency

Negative Binomial d.f.

	Parameters estimate		K-S test results	
	r	p	K-S value	p-value
A-set				
ET2	-	-	-	-
ET4	4.266	0.217	0.246	0.979
ET7	1.027	0.127	0.257	0.968
B-set				
ET2	5.212	0.796	0.298	0.899
ET4	3.853	0.224	0.230	0.989
ET7	0.536	0.086	0.268	0.953

Table 5.5: ET Frequency - Results of Negative Binomial Fitting

In general, for all ET data sets Poisson and Negative Binomial distributions provided good fit. The exception is ET2 A-set for Negative Binomial distribution. The fitting results can be also seen graphically in Figure 5.12 and Figure 5.13.

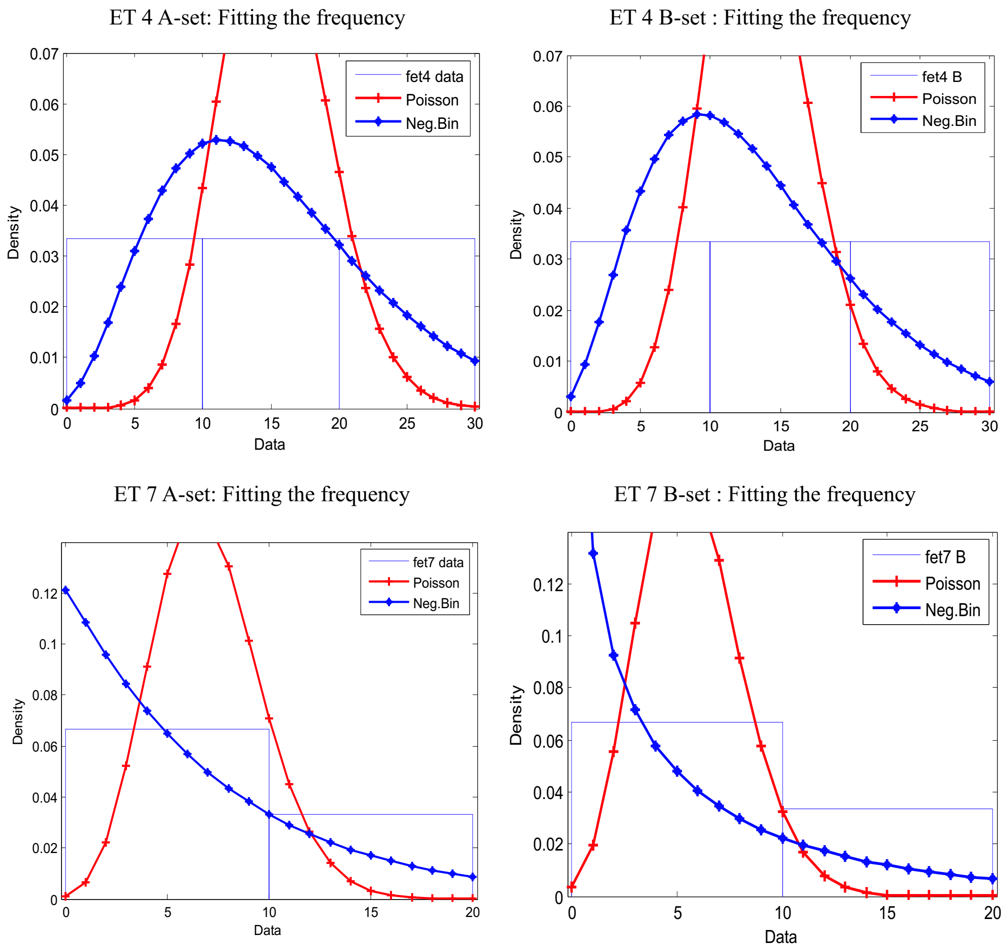

Figure 5.13: ET4 and ET7: Frequency

Event Types - Severity Fitting

The second part of fitting process deals with the severity of data. The results are presented in the same way as for the frequency of data, meaning that parameters estimates and K-S test results are given in the following tables and illustrated with graphics.

First we will start with the results from Weibull distribution fit to severity of ET data sets. See Table 5.6. Note that the K-S test values for A-set are smaller than the ones for B-set and that only for ET2 A-set p-value is not higher than the critical one.

Weibull d.f.

	Parameters estimate		K-S test results	
	a	b	K-S value	p-value
A set				
ET2	17,520	0.342	0.372	< 0.05
ET4	231,881	0.379	0.143	0.273
ET7	68,645	0.412	0.165	0.145
B set				
ET2	414,649	0.698	0.371	0.532
ET4	360,683	0.422	0.154	0.273
ET7	124,985	0.486	0.176	0.623

Table 5.6: ET Severity - Results of Weibull Fitting

LogNormal d.f.

	Parameters estimate		K-S test results	
	μ	σ	K-S value	p-value
A-set				
ET2	8.391	2.549	0.355	< 0.05
ET4	10.992	2.726	0.098	0.744
ET7	9.887	2.528	0.090	0.992
B-set				
ET2	11.951	2.633	0.390	0.467
ET4	11.604	2.318	0.141	0.376
ET7	10.714	1.991	0.147	0.824

Table 5.7: ET Severity - Results of LogNormal Fitting

The second distribution applied to the ET data sets is LogNormal distribution. Table 5.7 presents the fitting results which suggest good fit for all data sets apart from ET2 A-set. However, for ET2 B-set higher p-value is gained meaning that the data above higher threshold have been fitted better by LogNormal distribution than the one from A-set.

Fitting results for LogLogistic distribution are shown in Table 5.8. As expected all data sets are well fitted. For ET2 data set we have the same situation as for LogNormal distribution fit. Only in this case, according to K-S test, LogLogistic distribution has provided better fit than LogNormal.

LogLogistic d.f.

	Parameters estimate		K-S test results	
	μ	σ	K-S value	p-value
A set				
ET2	7.803	1.233	0.325	0.064
ET4	10.918	1.559	0.088	0.844
ET7	9.837	1.422	0.089	0.993
B set				
ET2	12.392	1.270	0.348	0.615
ET4	11.443	1.352	0.122	0.557
ET7	10.533	1.128	0.133	0.903

Table 5.8: ET Severity - Results of LogLogistic Fitting

Pareto d.f.

	Parameters estimate		K-S test results	
	x_m	k	K-S value	p-value
A-set				
ET2	1,000	0.674	0.533	<0.05
ET4	69	0.147	0.342	< 0.05
ET7	200	0.217	0.296	< 0.05
B-set				
ET2	3,200	0.257	0.462	0.266
ET4	3,056	0.279	0.185	0.114
ET7	4,277	0.425	0.128	0.923

Table 5.9: ET Severity - Results of Pareto Fitting

The fourth distribution adopted in fitting the severity of ET data sets is heavy-tailed Pareto distribution. The results are shown in Table 5.9. For all ET A-sets K-S test have rejected the hypothesis that they come from Pareto distribution since the K-S test values are smaller than critical ones. On the other side, for B-sets p-values are higher than 0.05, but only ET7 B-set provided good fit with p-value equal to 0.923.

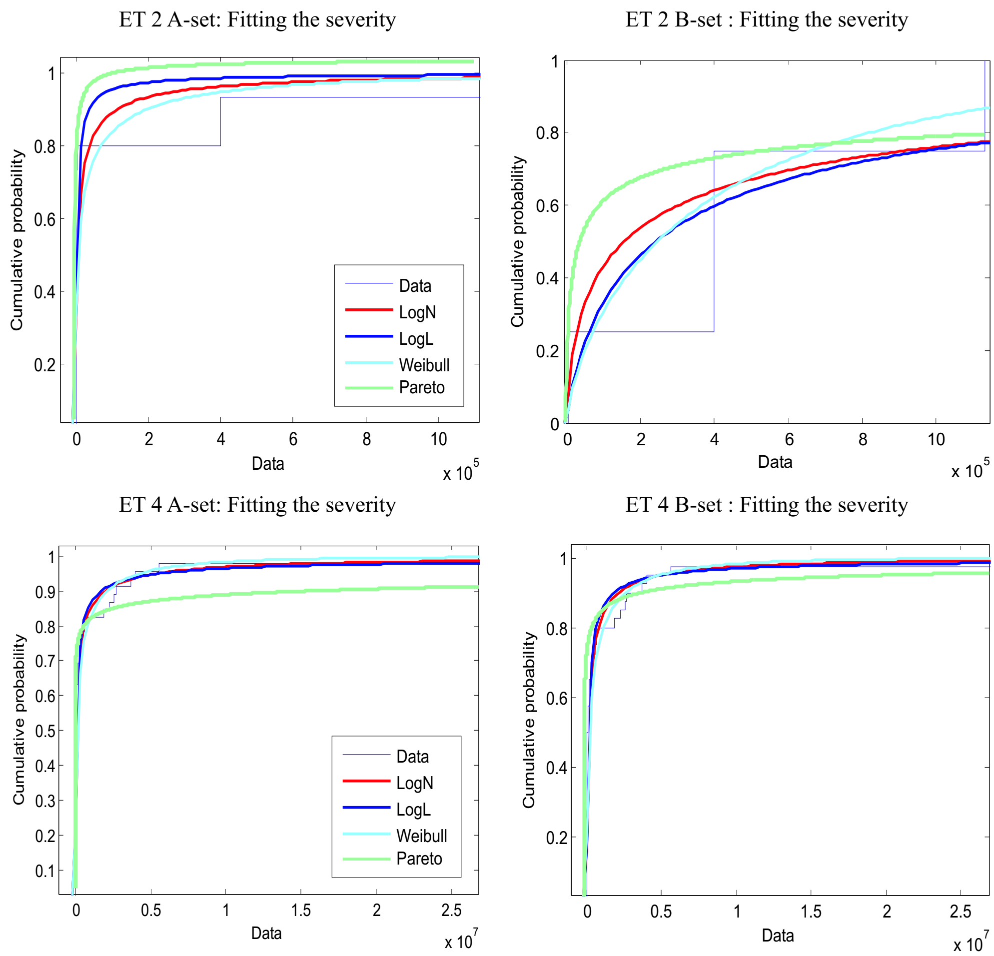

Figure 5.14: ET2 and ET4: Severity

The graphical presentation of fits for loss event type data sets are shown in Figure 5.14 and Figure 5.15.

However, the general conclusion from standard approach for loss event types can be summarized in the following way. In Table 5.10 and Table 5.11 the best choices of distributions according to K-S test are listed for both A-sets and B-sets data, respectively. The chosen distributions have the p-values varying from 0.5 to 0.99. Nevertheless, it should be noted non of considered distrtu did not that only for ET2 A-set data all considered distributions did not provide a good fit, apart from LogLogistic with the p-value equal to 0.064.

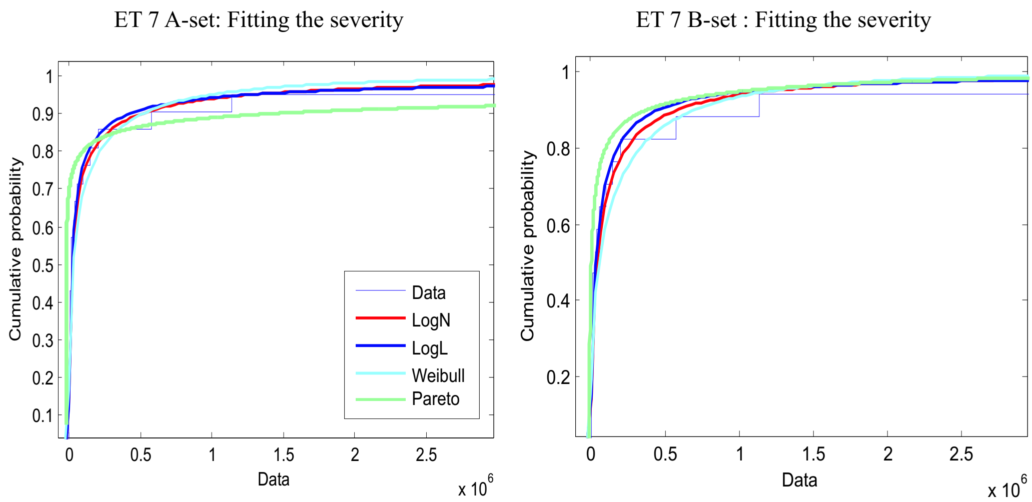

Figure 5.15: ET7: Severity

Standard Approach - MLE

	Frequency	Severity
A-set		
ET2	Poisson	LogLogistic
	5	7.803 1.233
	(0.931)	(0.064)
ET4	Neg.Bin	LogLogistic
	4.266 0.217	10.918 1.559
	(0.979)	(0.844)
ET7	Neg.Bin	LogLogistic
	1.027 0.127	9.837 1.422
	(0.968)	(0.993)

Table 5.10: ET A-set: Summary of Fitting Results

	Standard Approach - MLE	
	Frequency	Severity
B-set		
ET2	Poisson	LogLogistic
	1.333	12.392 1.270
	(0.923)	(0.615)
ET4	Neg.Bin	LogLogistic
	3.853 0.224	11.443 1.352
	(0.989)	(0.557)
ET7	Neg.Bin	Pareto
	0.536 0.086	4,277 0.425
	(0.953)	(0.923)
		LogLogistic
		10.533 1.128
		(0.903)

Table 5.11: ET B-set: Summary of Fitting Results

Business Lines - Frequency fitting

Now, we will consider the frequency of BL data sets. Generally, the fit results are quite good with the p-values range from 0.4 to 0.9. The exception are obviously the BL2 and BL4. The reported p-values for BL2 data set are quite low and close to critical one suggesting a poor fit of both distributions. Further, Negative Binomial distribution could not fit BL4 data while Poisson distribution has provided a very good fit with extremely high p-value of 0.97.

Poisson distribution function

	Parameter estimate	K-S test results	
	λ	K-S value	p-value
A-set			
BL2	4	0.648	0.091
BL3	10.333	0.475	0.379
BL4	9	0.254	0.971
BL5	4	0.428	0.510
B-set			
BL2	3.333	0.631	0.107
BL3	6	0.324	0.834
BL4	8.666	0.297	0.900
BL5	2.333	0.343	0.779

Table 5.12: BL Frequency - Results of Poisson Fitting

Negative Binomial d.f.

	Parameters estimate		K-S test results	
	r	p	K-S value	p-value
A set				
BL2	0.134	0.032	0.631	0.107
BL3	9.986	0.491	0.376	0.677
BL4	-	-	-	-
BL5	1.824	0.313	0.271	0.950
B set				
BL2	0.144	0.041	0.631	0.107
BL3	1.864	0.237	0.228	0.990
BL4	-	-	-	-
BL5	0.760	0.245	0.344	0.778

Table 5.13: BL Frequency - Results of Negative Binomial Fitting

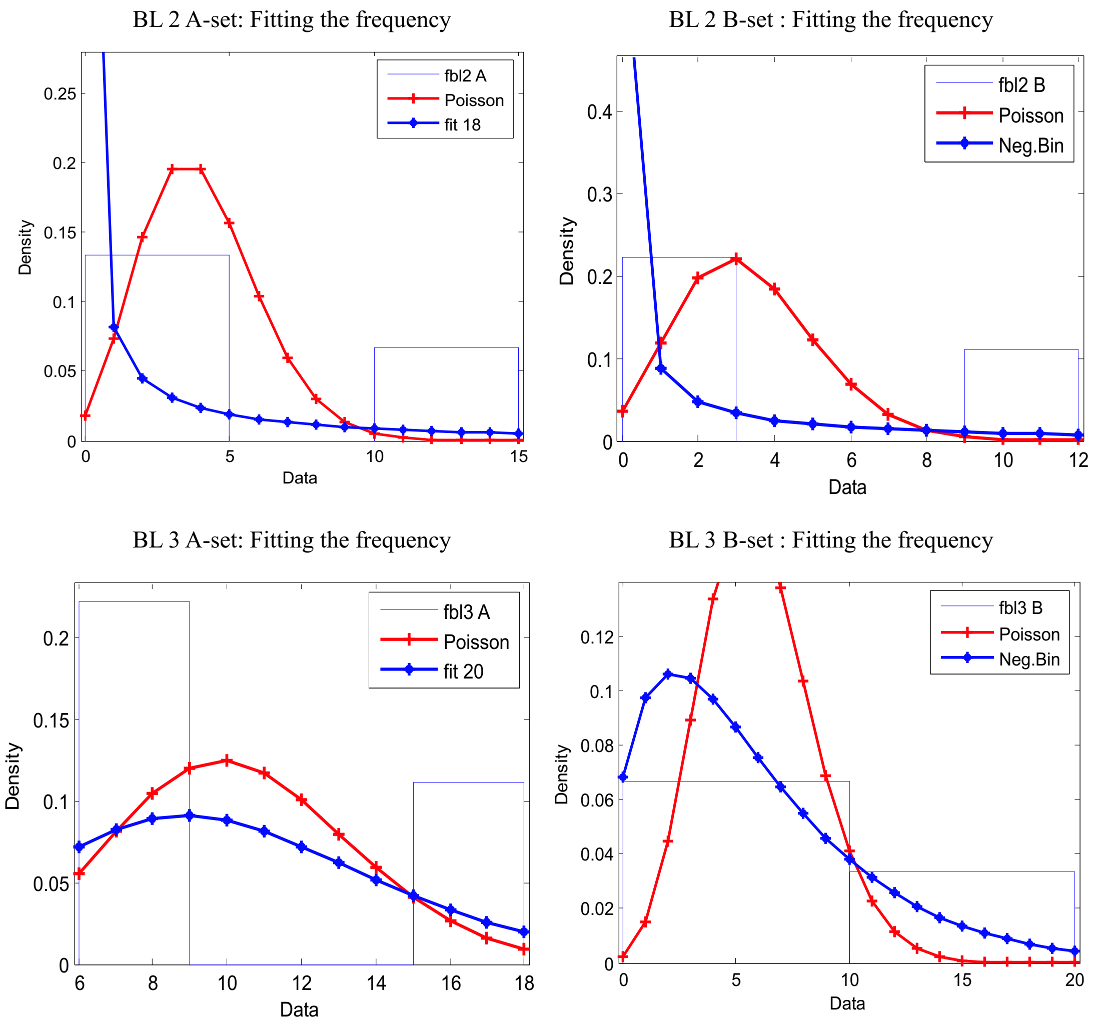

Figure 5.16: BL2 and BL3: Frequency

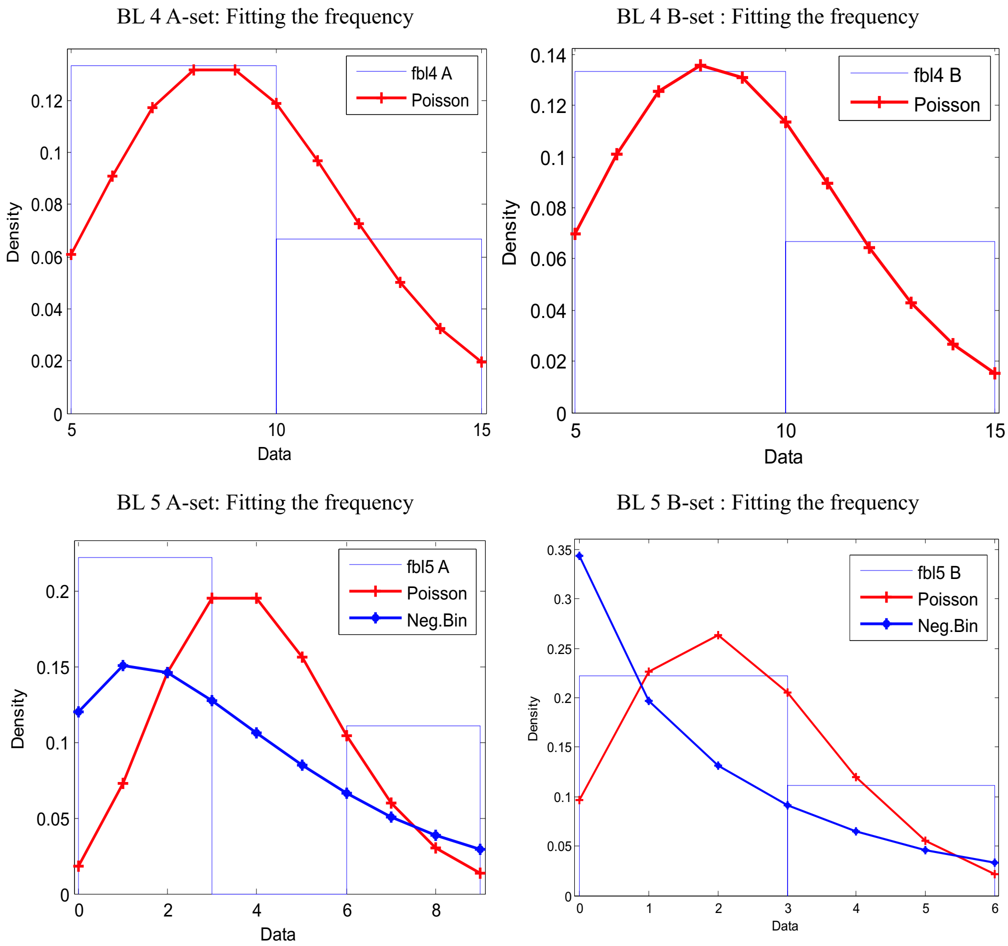

Figure 5.17: BL4 and BL5: Frequency

Figure 5.16 and Figure 5.17 show fitted distributions for all BL data sets.

Business Lines - Severity Fitting

Starting again from Weibull distribution, results of this severity fit are presented in Table 5.14. Immediately seen, Weibull distribution failed to fit all BL A-sets while for B-sets has provided a quite good fit.

The result of LogNormal fit are shown in Table 5.15. For all data set K-S test have accepted the hypothesis that LogNormal distribution can be used for their explanation.

Weibull d.f.

	Parameters estimate		K-S test results	
	a	b	K-S value	p-value
A set				
BL2	140,876	0.406	1	<0.05
BL3	19,021	0.434	1	<0.05
BL4	817,677	0.481	1	<0.05
BL5	13,404	0.643	1	<0.05
B set				
BL2	270,757	0.540	0.245	0.519
BL3	57,385	0.520	0.207	0.377
BL4	916,092	0.502	0.128	0.758
BL5	31,252	0.942	0.238	0.763

Table 5.14: BL Severity - Results of Weibull Fitting

LogNormal d.f.

	Parameters estimate		K-S test results	
	μ	σ	K-S value	p-value
A set				
BL2	10.453	3.118	0.196	0.697
BL3	8.757	2.069	0.146	0.489
BL4	12.515	2.248	0.100	0.937
BL5	8.693	1.630	0.168	0.851
B set				
BL2	11.551	1.931	0.217	0.678
BL3	10.016	1.804	0.144	0.818
BL4	12.689	2.098	0.105	0.920
BL5	9.770	1.236	0.235	0.774

Table 5.15: BL Severity - Results of LogNormal Fitting

The same conclusion can be obtained from the results of LogLogistic fit to BL data sets. If we compare the LogNormal and LogLogistic fit in terms of K-S test then LogLogistic distribution is more favorable for BL2, BL3 and BL5, while LogNormal distribution is for BL4.

Finally, the last used distribution is Pareto. As expected, BL A-sets were not fitted well by Pareto distribution, and according to K-S test only BL5 A-set and all B-sets have given a good test results. This is in line with the fact that data sets with a higher threshold and heavier tail properties are better fitted with heavy tailed distribution e.g. Pareto distribution.

LogLogistic d.f.

	Parameters estimate		K-S test results	
	μ	σ	K-S value	p-value
A set				
BL2	10.974	1.674	0.166	0.862
BL3	8.535	1.132	0.127	0.665
BL4	12.521	1.307	0.109	0.887
BL5	8.578	0.940	0.155	0.910
B set				
BL2	11.383	1.099	0.191	0.814
BL3	9.793	0.995	0.144	0.814
BL4	12.663	1.229	0.116	0.848
BL5	9.739	0.709	0.240	0.750

Table 5.16: BL Severity - Results of LogLogistic Fitting

Pareto d.f.

	Parameters estimate		K-S test results	
	x_m	k	K-S value	p-value
A-set				
BL2	69	0.161	0.394	< 0.05
BL3	200	0.289	0.307	< 0.05
BL4	2,294	0.221	0.258	< 0.05
BL5	1,000	0.560	0.250	0.387
B-set				
BL2	11,589	0.455	0.154	0.956
BL3	3,056	0.502	0.144	0.820
BL4	9,123	0.280	0.202	0.210
BL5	4,277	0.709	0.223	0.831

Table 5.17: BL Severity - Results of Pareto Fitting

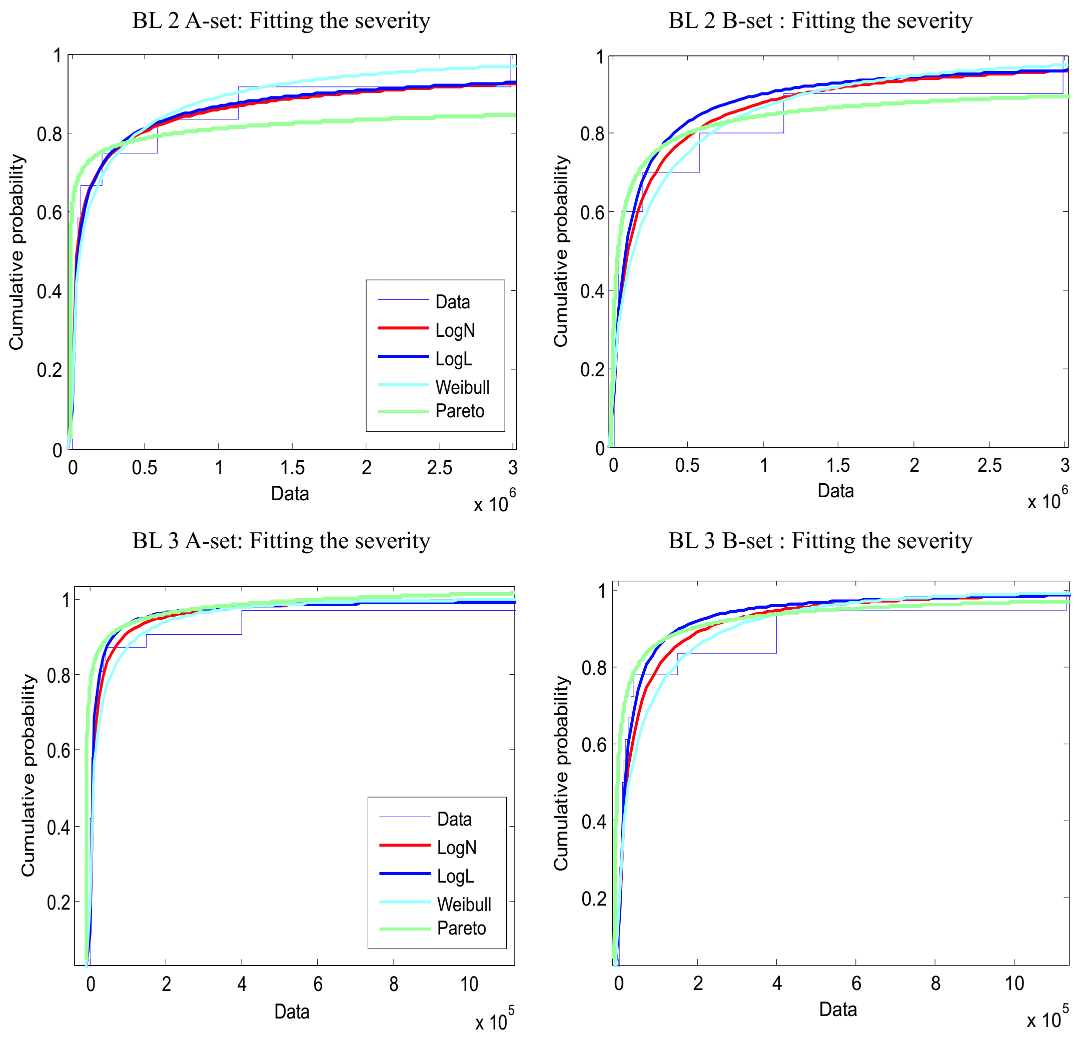

Figure 5.18: BL2 and BL3: Severity

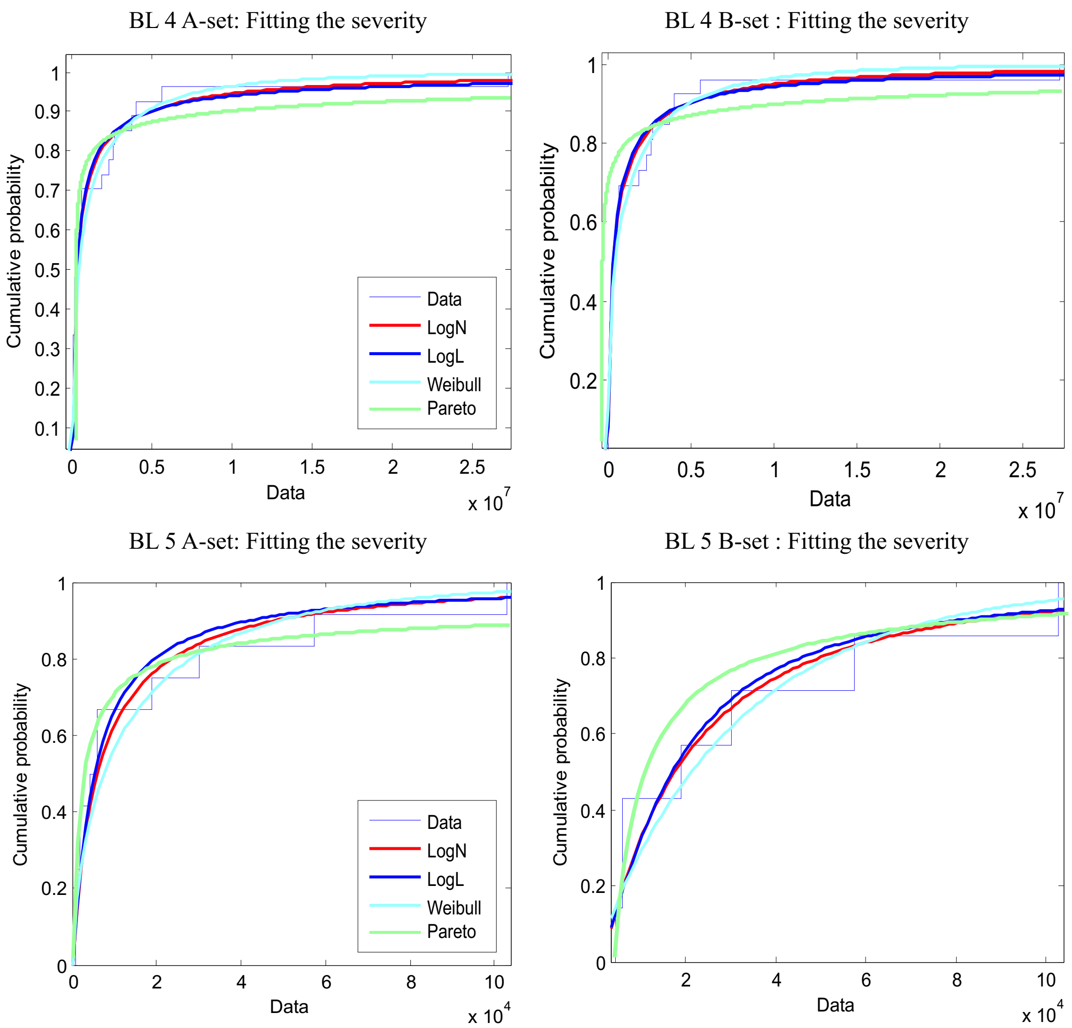

Figure 5.19: BL4 and BL5: Severity

The graphical presentation of fitting the severity of BL data set is shown in Figure 5.18 and Figure 5.19.

The review of the best choices of distributions for business lines is given in Table 5.18 and Table 5.19. In the case where two distributions are listed also the second best choices are provided.

Standard Approach - MLE

	Frequency	Severity
A-set		
BL2	Neg.Bin	LogLogistic
	0.134 0.032	10.974 1.674
	(0.107)	(0.862)
BL3	Neg.Bin	LogLogistic
	9.986 0.491	8.535 1.132
	(0.677)	(0.665)
BL4	Poisson	LogNormal
	9	12.515 2.248
	(0.971)	(0.937)
BL5	Neg.Bin	LogLogistic
	1.824 0.313	8.578 0.940
	(0.950)	(0.910)

Table 5.18: BL A-set: Summary of Fitting Results

It can be conclude that as far as the A-set data is considered the LogLogistic distribution is the most common one. This holds for both event types and business lines. On the other hand, for business lines B-set data the most appropriate distribution is Pareto according to K-S test. For loss event types B-set data it is LogLogistic.

Although we can select the most used distribution in OR analysis and draw some conclusions, it is important to note that there is no agrement and rules which distributions will be more suitable for some other data set. That is why every data set must be considered and fitted separately.

Standard Approach - MLE

	Frequency	Severity
B-set		
BL2	Poisson 3.333 (0.107)	Pareto 11,589 0.455 (0.956) LogLogistic 11.838 1.099 (0.814)
BL3	Neg.Bin 1.864 0.237 (0.677)	Pareto 3,056 0.502 (0.820) LogNormal 10.016 1.804 (0.818)
BL4	Poisson 8.666 (0.900)	LogNormal 12.689 2.098 (0.920)
BL5	Poisson 2.7333 (0.779)	Pareto 4,277 0.709 (0.831) LogNormal 9.770 1.236 (0.774)

Table 5.19: BL B-set: Summary of Fitting Results

5.2.2 Adjusted Approach

Results Obtained Using Numerical Optimization

1. Severity fitting

The Adjusted approach is performed according to the given explanation in Section 4. In order to solve the maximization problem in Equation 4.12 we have used the numerical optimization tools. Naturally, since the parameters for frequency depend on the estimation of severity parameters, we have first computed the severity parameters and latter frequency parameters according to Equation 4.13.

It should be stressed that this adjusted approach is performed only on the B-set for both of ET and BL. This is due to the fact that A-sets are considered as complete data sets while B-sets with higher threshold have property of left-truncated sets and lack of information. Certainly, under this approach the adjusted Kolmogorov-Smirnov goodness-of-fit test denoted by K-S* is performed.

We will start with the results for Weibull distribution in Table 5.20. They imply that Weibull distribution does not provide a good fit since the p-values are not much higher than $\alpha = 0.05$. Only ET2 B-set is fitted well with p-value equal to 0.42. If we want to give a general conclusion on usage of Weibull distribution for fitting the severity of OR losses then there are facts against it. Namely, Weibull distribution, according to K-S test and considering our data sample does not provide a good fit, and thus, in vast majority of cases it can not be considered as a good choice for severity distribution. Moreover, this conclusion holds for results under standard and adjusted approach.

Adjusted approach: Weibull d.f.

	Parameters estimate		K-S test results	
	a	b	K-S* test	p-value
ET B-set				
ET2	414,645	0.623	0.564	0.420
ET4	360,686	0.353	2.035	< 0.05
ET7	124,992	0.401	1.515	< 0.05
BL B-set				
BL2	270,753	0.465	1.072	0.101
BL3	58,377	0.423	1.781	0.050
BL4	916,082	0.452	0.922	0.371
BL5	30,258	0.774	1.071	0.100

Table 5.20: Results of Weibull Fit under Adjusted Approach

Adjusted approach: LogNormal d.f.

	Parameters estimate		K-S test results	
	μ	σ	K-S* test	p-value
ET B-set				
ET2	11.387	2.725	0.651	0.370
ET4	10.601	2.975	1.423	0.251
ET7	8.979	2.904	1.526	0.113
BL B-set				
BL2	11.321	2.043	0.693	0.312
BL3	2.167	4.342	3.851	0.184
BL4	12.598	2.159	0.495	0.767
BL5	9.206	1.518	0.937	0.265

Table 5.21: Results of LogNormal fit under Adjusted Approach

The second fitted distribution is LogNormal. The results shown in Table 5.21 suggest better fit than the one for Weibull distribution. Yet the p-values are not much higher than α, apart from the BL4 where p-value is 0.767.

Adjusted approach: LogLogistic d.f.

	Parameters estimate		K-S test results	
	μ	σ	K-S* test	p-value
ET B-set				
ET2	12.025	1.508	0.592	0.335
ET4	10.633	1.741	1.363	0.291
ET7	9.315	1.599	1.348	0.159
BL B-set				
BL2	11.133	1.262	0.653	0.412
BL3	5.446	1.813	3.429	< 0.05
BL4	12.537	1.327	0.562	0.670
BL5	9.162	0.944	0.955	0.262

Table 5.22: Results of LogLogistic Fit under Adjusted Approach

The results of fitting the LogLogistic distribution to the B-sets using the adjusted approach presented in Table 5.22 are quite similar to the results of LogNormal distribution fit. Going further, the forth considered distribution is Pareto and according to the p-values almost all B-sets have shown a good fit with exception of ET4 where p-value is close to α.

Adjusted approach: Pareto d.f.

	Parameters estimate		K-S test results	
	x_m	k	K-S* test	p-value
ET B-set				
ET2	3,200	0.254	0.443	0.881
ET4	3,056	0.278	1.045	0.150
ET7	4,277	0.369	0.815	0.892
BL B-set				
BL2	11,589	0.282	2.007	0.786
BL3	3,056	0.497	0.540	0.610
BL4	9,123	0.214	2.437	0.665
BL5	4,277	0.567	0.608	0.891

Table 5.23: Results of Pareto Fit under Adjusted Approach

2. Frequency fitting

Having computed the severity parameters the calculation of frequency parameters is quite simple. First we need to decide which severity distribution function is the most appropriate according to the K-S* test and then compute the "information loss" $F(u;\hat{\theta}^{adj})$. The second step is to divide already estimated MLE frequency parameters by probability that loss is greater than threshold u, i.e.

$$\hat{\tau}^{adj} = \frac{\hat{\tau}^{mle}}{1 - F(u;\hat{\theta}^{adj})} .$$

Since the cumulative distribution function is from the range [0,1] we expect that the adjusted frequency parameters will be higher than the one estimated by MLE method. In this way we have "corrected" the estimated parameters by the information that all given and recorded losses come from area $x \geq u$.

In the following table the new estimated frequency parameters are listed. The best choice of frequency distributions is done by the means of results in standard approach, and the selected severity distributions are marked in the table.

Adjusted Approach: Frequency

	Poisson	Neg. Binomial		K-S test results	
	λ	p	r	K-S value	p-value
ET B-set					
ET2 Pareto (0%)	1.333			0.286	0.923
ET4 LogL (18.11%)		4.706	0.274	0.264	0.959
ET7 Pareto (0%)		0.536	0.086	0.268	0.953
BL B-set					
BL2 Pareto (0%)	3.333			0.631	0.107
BL2 LogL (7.75%)	3.613			0.639	0.099
BL3 Pareto (0%)		1.864	0.237	0.228	0.990
BL4 LogN (1.67%)	8.813			0.278	0.938
BL5 Pareto (0%)	2.333			0.343	0.779

Table 5.24: Adjusted Frequency Distribution Results

In the Table 5.24 it is also reported the percentage of "information loss" as the value of $F(u; \hat{\theta}^{adj})$ in the brackets. In the case of Pareto severity distribution there is no changes in values of frequency parameters since the cumulative Pareto distribution for the threshold ($\min(x)$) is equal to zero. That is the property of Pareto distribution and it can also be seen in Figure 4.5.

The other adjustments are not so severe meaning the difference is around 10^{-2}. This is on the account of the fact that our threshold u is not a big number. For the bigger threshold the bigger difference in values is expected.

Results of EM algorithm

The EM algorithm for LogNormal-Poisson aggregated loss distribution explained in Section 4 was applied on the considered data. This is an alternative way of including the effect of missing data in parameters estimation. Naturally, the EM algorithm was applied only on B-set data and the K-S* test is performed for LogNormal severity distribution. The results are given in Table 5.25.

According to K-S* test for ET2 and BL4 data set the p-values are 0.721 and 0.810, respectively, suggesting a good fits.

If we compare these results for severity parameters μ and σ with the one from Standard approach (Table 5.7 and Table 5.13) we can see that there is a difference between estimated parameters. Namely, the difference is around 10^{-2}.

EM algorithm: LogNormal-Poisson d.f.

	Poisson	LogNormal		K-S test results	
	λ	μ	σ	K-S* test	p-value
ET B-set					
ET2	1.352	11.948	2.283	0.557	0.721
ET4	13.599	11.599	2.294	1.334	0.145
ET7	5.825	10.706	1.937	1.095	0.278
BL B-set					
BL2	3.363	11.551	1.833	0.792	0.574
BL3	6.271	9.993	1.764	1.238	0.144
BL4	8.699	12.689	2.058	0.572	0.810
BL5	2.382	9.767	1.146	0.907	0.212

Table 5.25: EM for LogNormal-Poisson d.f.

On the other side, the frequency parameter λ calculated under the EM algorithm is higher than the one from MLE estimation (Table 5.4 and Table 5.12). Obviously, if the effect of left-truncated data is included, the number of loss events is expected to be higher.

Summary of Adjusted Approach

Adjusted Approach

	Frequency	Severity
ET B-set		
ET2	Poisson	Pareto
	1.333	3,200 0.254
	(0.923)	(0.881)
ET2 EM	Poisson	LogNormal
	1.352	11.948 2.283
	(0.927)	(0.721)
ET4	Neg. Binomial	LogLogistic
	4.706 0.274	10.633 1.741
	(0.959)	(0.291)
ET7	Neg. Binomial	Pareto
	0.536 0.086	4,277 0.369
	(0.953)	(0.892)

Table 5.26: ET: Summary of Adjusted Approach

Adjusted Approach	Frequency	Severity
BL B-set		
BL2	Poisson	Pareto
	3.333	11,589 0.282
	(0.107)	(0.786)
	Poisson	LogLogistic
	3.613	11.133 1.262
	(0.099)	(0.412)
BL3	Neg. Binomial	Pareto
	1.864 0.237	3,056 0.497
	(0.990)	(0.610)
BL4	Poisson	LogNormal
	8.813	12.598 2.159
	(0.779)	(0.767)
BL4 EM	Poisson	LogNormal
	8.699	12.689 2.058
	(0.909)	(0.810)
BL5	Poisson	Pareto
	2.333	4,277 0.567
	(0.799)	(0.891)

Table 5.27: BL: Summary of Adjusted Approach

5.2.3 EVT Analysis

EVT theory has two methods Black Maxima and Peak Over Threshold (POT). In our case, there was no point in applying the the first method. The main reason lays in the fact that it is really hard to find at least 5 maximums for considering time intervals. Namely, if we divide our time horizon in time intervals (e.g. months) it is more probably that there will be lack of data (no observations) in the vast majority of intervals. Also, if we take years as time intervals then only 3 data points will be gained. Consequently, this method was excluded from analysis and only Peak Over Threshold method was conducted, even though the number of excess the threshold is sometimes less than 10.

The EVT analysis was performed only on the B-set data meaning that all data sets have the threshold 3,000. The $GPD(x; \xi, \beta)$ distribution fitted to the data was explained before by Equation 4.39. The estimated parameters ξ and β are reported in the following table.

GPD

	number of excess	Parameters estimate	
	n	ξ	β
ET2	4	-2.131	848,856
ET4	40	2.064	47,381
ET7	17	1.629	24,912
BL2	10	1.658	62,l532
BL3	18	1.439	13,145
BL4	26	1.594	213,704
BL5	7	0.224	24,714

Table 5.28: Results of GDP Fit

If we put our attention on estimated parameter ξ it can be concluded that all data sets apart from ET2 have a heavy tail property since it holds $\xi > 0$. Of course, this is in line with all previous findings. Yet, it should be noted that the analysis for ET2 and BL5 with only 4 and 7 excesses can not give us the reliable results. Obviously, the number of excesses can be limiting factor.

In Figure 5.20 and Figure 5.21 it can be seen how well GDP has fitted the data sets.

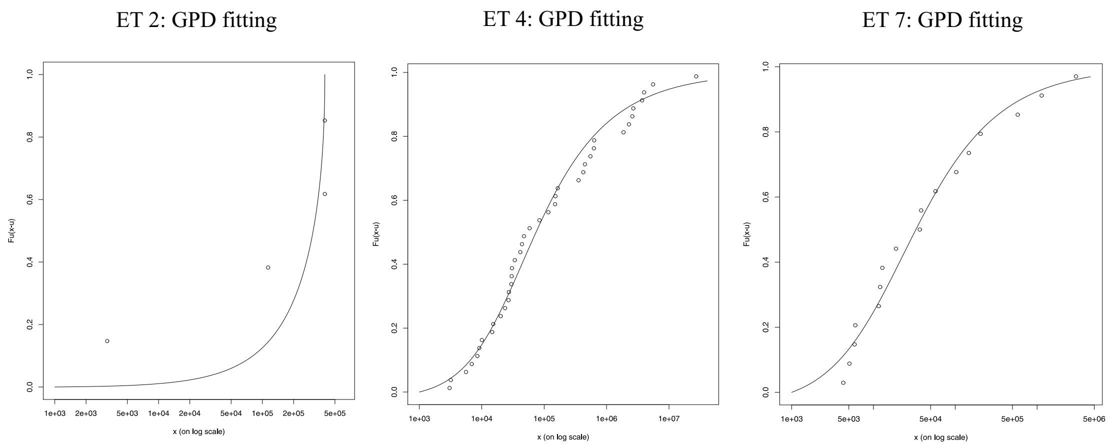

Figure 5.20: GDP

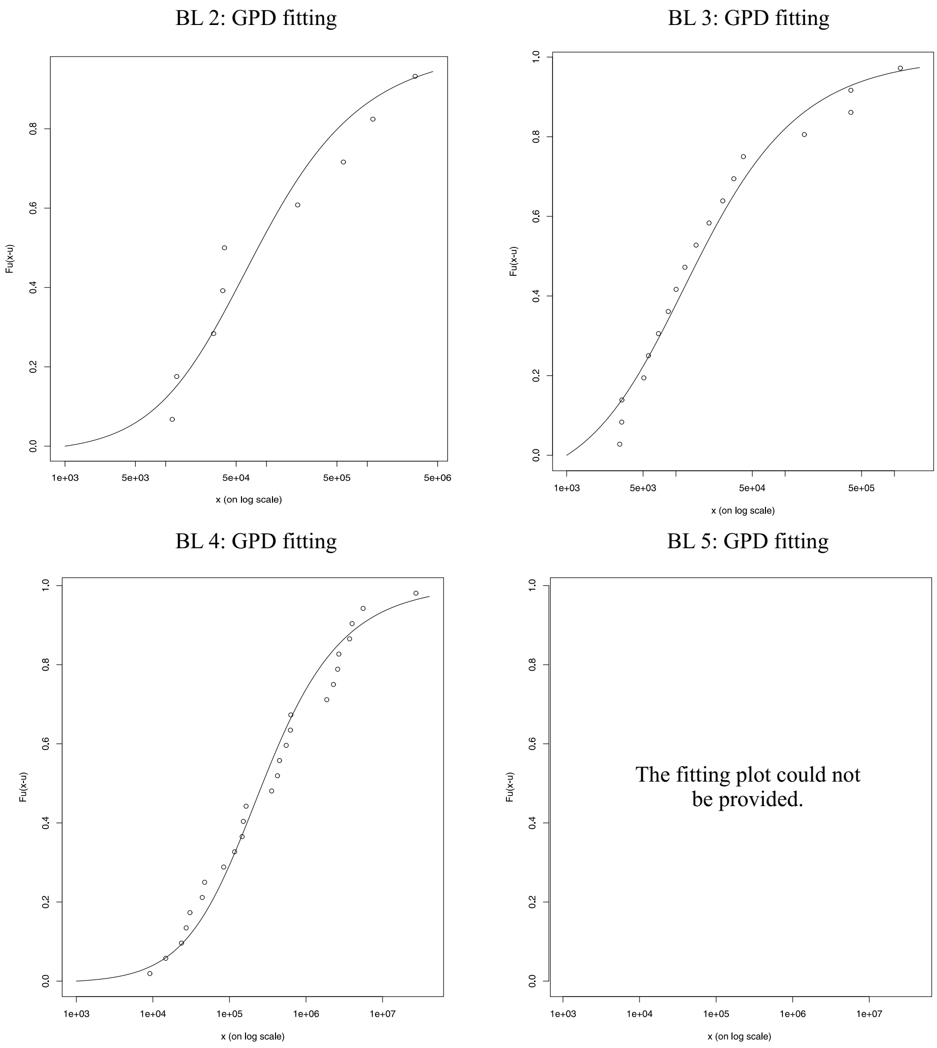

Figure 5.21: GDP

5.3 Monte Carlo Simulation

Having selected the best fitted distributions for both of frequency and severity of loss data sets the process of theirs aggregation is performed via Monte Carlo simulation.

It should be noted that the results from OR VaR Monte Carlo simulation for 99% and 99.9% level are omitted from the paper. Namely, the results for these high percentiles can not be considered as transparent measures since we have start the analysis with very small number of observations, and thus, the high precision could not be obtained. Following the reasoning, the reported results are for the levels 90% and 95%.

5.3.1 Standard Approach Results

OR VaR (in 000)

	90%	95%
ET A-set		
ET2	340	773
ET4	152,535	435,534
ET7	8,876	22,973
ET B-set		
ET2	6,217	15,749
ET4	80,399	194,840
ET7 PR	48,021	270,144
ET7 LL⋆	4,934	10,363

Table 5.29: ET- OR VaR

OR VaR (in 000)

	90%	95%
BL A-set		
BL2	18,782	85,898
BL3	1,203	2,428
BL4	61,120	101,377
BL5	234	418
BL B-set		
BL2 PR	24,258	113,999
BL2 LL⋆	4,673	9,660
BL3 PR	10,451	42,261
BL3 LN⋆	1,645	2,508
BL4 LN	51,886	82,636
BL5 PR	400	1,022
BL5 LN⋆	208	293

Table 5.30: BL -OR VaR

The mark ⋆ in the above tables suggests that the second best choice for severity distribution function was considered. The idea was to see the difference between OR VaR values if the heavy tailed Pareto or medium tailed LogNormal and LogLogistic distribution is selected. Obviously, the difference in some cases can be quite big, up to 10^2 at the level 95% and even more for higher levels. This findings are in the line with the fact that Pareto distribution puts more weights to tail of data sets.

On the other side, the question is does the Kolmogorov-Smirnov test provide the appropriate choice of fitted distribution function. Namely, test calculates the maximum difference between fitted and empirical distribution without considering where the difference has occurred. Perhaps some other goodness of fit tests such as Anderson-Darling, Cramer von Mises, would lead to other conclusions. These issues are left for further investigations.

5.3.2 Adjusted Approach Results

OR VaR (in 000)

	90%	95%
ET B-set		
ET2 PR	72,022	1,193,310
ET2 EM⋆	4,659	10,217
ET4 LL	187,155	618,978
ET7 PR	185,425	1,139,170
BL B-set		
BL2 PR	2,513,882	32,379,317
BL2 LL⋆	8,785	20,711
BL3 PR	11,227	45,847
BL4 LN	54,425	87,992
BL4 EM⋆	48,073	75,169
BL5 PR	1,136	3,835

Table 5.31: Adj. App. - OR VaR

Table 5.31 gives results of Monte Carlo simulation for adjusted approach. Firstly, as it was expected the OR VaR values under adjusted approach are much higher than the ones calculated under standard approach. The influence of the threshold and missing values is apparently important since it considerably changes the OR VaR value. Thus, OR risk management should record as much as possible of loss events leaving the number of missing data at reasonable low level. These issues are investigated in more details in papers [14], [15] and [16].

Secondly, the empirical results suggest that severity Pareto distribution gives much higher OR VaR values than some other medium-tailed distributions.

Apparently, in some cases the deference can be more than 10^3 for 95% level and up to 10^6 for higher confidence levels. Knowing that our analysis was started with very small number of data, it seems reasonable not to consider the Pareto distribution for BL2 and ET2 data set under 95% level. Obviously, in these particular cases the difference in OR VaR values is quite big and more probably driven from limiting number of data then from the fact that Pareto distribution provided better fit.

Further, in Table 5.31 we have results for BL4 B-set from numerical optimization and EM algorithm for Poisson-LogNormal aggregated distribution. Certainly, the reported difference among OR VaR values comes from the difference in the estimated parameters μ and σ. If we go back to Table 5.27 the results under EM algorithm are higher for μ and smaller for σ. Also, the frequency parameter λ is smaller suggesting less number of simulated events. This explains the difference in OR VaR values.

5.4 OR Capital Charge

In this section the final estimated values for OR capital charge for the whole bank are given. The results are arranged according to performed ET or BL analysis. Once again the mark $\star$ denotes that the second best choice for severity distribution is used.

Capital Charge - Standard App.		
	90%	95%
	ET A-set	
CC1	161,750,593	459,280,002
CC2	18,544,674	32,915,159
	ET B-set	
CC1	134,637,142	480,732,729
CC2	23,025,246	62,188,531
	ET B-set $\star$	
CC1	91,549,989	220,951,592
CC2	14,939,319	24,338,676

Table 5.32: ET Capital Charge

CC1 is the capital charge calculated under the assumption of perfect correlation among event types or business lines. The formula is simply the sum of the corresponding OR VaR values.

CC2 is the capital charge computed when the correlation effect is included according to following equation

$$\mathrm{CC2}_\alpha = \sum_m \mathrm{EL}_m + \sqrt{\sum_{m,n \ m \neq n} k_{m,n}(\mathrm{CC}_m - \mathrm{EL}_m)(\mathrm{CC}_n - \mathrm{EL}_n)} \qquad (5.1)$$

In our case, the correlation factor k is set to be 5% for all data sets, and for expected loss the median of sample is used. The main reason for using the median measure instead of mean value is the high kurtosis and skewness values of data sets and also the small number of data. These properties of data sets lead to big difference between mean and median values. In our particular case the mean value is sometimes higher than the 90^{th} quantile. Accordingly, we have decided to use median as expected loss measure.

Capital Charge - Standard App.		
	90%	95%
	BL A-set	
CC1	81,428,228	190,120,649
CC2	28,845,949	48,792,425
	BL B-set	
CC1	86,995,636	239,917,561
CC2	29,209,043	54,233,210
	BL B-set ⋆	
CC1	58,412,075	95,096,664
CC2	24,189,726	32,321,376

Table 5.33: BL Capital Charge

Capital Charge - Adjusted App.		
	90%	95%
	ET B-set ⋆	
CC1	377,243,287	1,998,365,837
CC2	50,372,306	215,555,354
	BL B-set ⋆	
CC1	69,220,079	145,561,941
CC2	25,738,589	39,795,831

Table 5.34: ET and BL Capital Charge

In Table 5.32 and Table 5.33 the computed capital charges for ET and BL analysis are listed, respectively. Immediately seen, for both analysis CC2 values are smaller than CC1. This implies that including the correlation effect leads to reduction of the capital charge.

Also, as expected the results from adjusted approach in Table 5.34 suggest a higher capital charge than the one under standard approach. Further, the values CC1 are still less than CC2.

5.4.1 Summary of OR VaR Estimated Values

ET analysis - Capital Charge

	90%	95%
Standard App: ET B-set ⋆		**CC2**
MIN	14,939,319	24,338,676
in eur	186,741	304,233
Adjusted App: ET B-set ⋆		**CC1**
MAX	377,243,287	1,998,365,837
in eur	4,715,541	24,979,572

Table 5.35: Summary of ET Analysis

BL analysis - Capital Charge

	90%	95%
Standard App: BL B-set ⋆		**CC2**
MIN	24,189,726	32,321,376
in eur	302,372	404,017
Standard App: BL B-set		**CC1**
MAX	86,995,636	239,917,561
in eur	1,087,445	2,998,969

Table 5.36: Summary of BL Analysis

The above Tables 5.35 and 5.36 give the minimum and maximum capital charges that were obtained from diverse analysis. Apparently, the differences are huge, up to 10^8. The minimum capital charges are obtained from second best choices of severity distributions, and maximum from adjusted approach and Pareto distribution. It is evident that the medium-tailed distributions LogNormal and LogLogistic provided smaller capital charges when correlation effect was included.

All this suggests that risk managers should take the measuring of operational risk with more caution since the final values are very sensitive to the choices of distributions, parameters estimation methods, recorded threshold, correlation effect ect. Indeed, a very little difference in estimated parameters can produces a great difference in final capital charge figure. Also, according to our empirical study based on provided historical data the effect of correlation is not irrelevant as well as the choice of heavy-tailed or medium-tailed severity distribution function.

Bibliography

[1] Basel Committee on Banking Supervision, *Working Paper on the Regulatory Treatment of Operational Risk*, September 2001

[2] Basel Committee on Banking Supervision, *Sound Practices for the Management and Supervision of Operational Risk*, February 2003

[3] Basel Committee on Banking Supervision, *International Convergence of Capital Measurement and Capital Standards - A Revised Framework Comprehensive Version*, June 2006

[4] Official site of Bank of International Settlement, www.bis.org

[5] P. de Fontnouvelle, V. DeJesus-Rueff, J. Jordan and E. Rosengren, *Capital and Risk: New Evidence on Implications of Large Operational Losses*, Federal Reserve Bank of Boston, September 2003

[6] P. de Fontnouvelle and E. Rosengren, *Implications of Alternative Operational Risk Modeling Techniques*, Federal Reserve Bank of Boston, John Jordan FitchRisk, June 2004

[7] P. de Fontnouvelle, V. DeJesus-Rueff, J. Jordan, E. Rosengen, *Using Loss Data to Quantify Operational Risk*, Federal Reserve Bank of Boston, John Jordan FitchRisk, Aprile 2003

[8] A.P. Dempster, N.M. Laird and D.B.Rubin, *Maximum Likelihood from Incomplete Data via the EM Algorithm*, Journal of the Royal Statistical Society, Series B (Methodological), 39(1), 138, 1977

[9] P. Embrechts, C. Klppelberg and T. Mikosch, *Modeling Extremal Events for Insurance and Finance*, Berlin Springer-Verlag, 1997.

[10] A. Frachot, P. Georges and T. Poncalli, *Loss Distribution Approach for operational risk*, Credit Lyonnais, Groupe de Recherche Operationnelle, Working Paper, 2001.

[11] N. Baud, A. Frechot, T. Roncalli, *How to Avoid Over-estimating Capital Charge for Operational Risk?*, Credit Lyonnais, Groupe de Recherche Operationnelle, Working Paper, 2002.

[12] A. Frachot, T. Roncalli and E. Salomon, *The Correlation Problem in Operational Risk*, Groupe de Recherche Operationnelle Group Risk Management, Credit Agricole SA, France, 2004.

[13] A. Chernobai, C. Menn, S. Rachev, S. Truck, *Estimation of Operational Value-at-Risk in the Presence of Minimum Collection Thresholds*, September 2005.

[14] A. Cheronbai, S. Menn, S. Truck, S. Rachev, *A note on the estimation of the frequency and severity distribution of operational losses*, Mathematical Scientist 30 (2), 2005.

[15] A.Chernobai, C. Menn, S. Rachev, S. Truck, M. Moscadelli, *Treatment of Incomplete Data in the Field of Operational Risk: The Effects on Parameter Estimates, EL and UL Figures*, University of California, 2006.

[16] A. Chernobai, K. Burnecki, S. Rachev, S. Truck, R. Weron, *Modelling catastrophe claimes with left-truncated severity distributions*, Computational Statistics, 21(3), 2006.

[17] R. Giacometti, S. Rachev, A. Chernobai, M. Bertocchi and G. Consigli, *Practical Operational Risk*, JOR, 2007

[18] R. Fisher and L. Tippet, *Limiting Forms of the Freuqency Distribution of the Largest or Smallest Member of a Sample*, Proceedings of the Cambridge Philosophical Society, *24* , 1928.

[19] M. Moscadelli, *The Modelling of Operational Risk: Experience with the Analysis of the Data Collected by the Basel Committee*, Bank of Italy, working paper, 2004.

[20] A. Chernobai, S. Rachev, F. Fabozzi, *Composite Goodness-of-Fit Tests for Left-Truncated Loss Samples*, Technical report, University of California Santa Barbara, 2005.

[21] Marcelo G. Cruz, *Modeling, Measuring and Hedging Operational Risk*, John Wiley & Sons, 2002

[22] Carol Alexander, *Operational Risk- Regulation, Analysis and Management*, Finantial Times Prentice Hall, 2003

[23] Micheal Frenkel, Urlich Hommel, Markus Rudolf (Editors) 2nd Edition, *Risk MAnagement- Challenge and Opportunity*, Springer, 2005

[24] Philippe Jorion, *Value at Risk- The New Benchmark for Managing Financial Risk* 2nd Edition, McGraw-Hill, 2001

[25] Elena A. Medova, *Measuring risk by Extreme Values*, Risk, November 2000

[26] Kwabena Adusei-Poku, *Operational Risk management - Implementing a Bayesian Network for Foreign Exchange and Money Market Settlement*, PhD Dissertation at the Faculty of Economics and Business Administration of the University of Göttingen, 2005

[27] Martin Sköld, *Computer Intensive Statistical Methods*, Lund University, Centre for Mathematical Sciences Mathematical Statistics, Lecture notes for FMS091/MAS221 October 2005, 2nd printing August 2006

UNIVERZITET U NOVOM SADU
PRIRODNO-MATEMATIČKI FAKULTET
KLJUČNA DOKUMENTACIJSKA INFORMACIJA

Redni broj:

RBR

Identifikacioni broj:

IBR

Tip dokumentacije: Monografska dokumentacija

TD

Tip zapisa: Tekstualni štampani materijal

TZ

Vrsta rada: Magistarska teza

VR

Autor: Ivana Manić

AU

Mentor: prof. dr Nataša Krejić

MN

Naslov rada: Procena i upravljanje operativnim rizikom pomoć u matematičkih modela

NR

Jezik publikacije: Engleski

JP

Jezik izvoda:

JI

Zemlja publikovanja: Srbija

ZP

Uže geografsko područje: Vojvodina

UGP

Godina: 2007.

GO

Izdavač: Autorski reprint

IZ

Mesto i adresa: Novi Sad, Prirodno-matematički fakultet, Trg Dositeja Obradovića 4

MA

Fizički opis rada:(5/111/32/39/11/21/0)
(broj poglavlja/strana/lit. citata/tabela/slika/grafika/priloga)

FO

Naučna oblast: Matematika

NO

Naučna disciplina: Primenjena matematika

ND

Ključne reči: Matematički modeli u ekonomiji, upravljanje rizikom, operativni rizik, Value at Risk, Loss Distribution Approach, EM algoritam, EVT teorija, Bazel II

PO

UDK:

Čuva se: Biblioteka Departmana za matematiku i informatiku, Novi Sad

ČU

Važna napomena:

VN

Izvod: Magistarska teza je posvećena matematičkim modelima za odredjivanje operativnog rizika u finansijskim institucijama. Izračunavanje operativnog rizika je relativno novi problem u oblasti upravljanja finansijskim rizicima. U okviru teze tri osnovna pristupa su objašnjena Osnovni, Standardni i Napredni pristup. Akcenat u radu je na metodi Raspodele gubitaka koja se bazira na različitim matematičkim modelima kao što su metod maksimalne verodostojnosti, fitovanje podataka raspodelama debelog repa, algoritam Expectation-Maximization, teorija ekstremnih vrednosti, Monte Carlo simulacija, statističkim testovima itd. Rad sadrži i epirijske rezultate posmatranih metoda.

IZ

Datum prihvatanja teme od NN veća: 20.03.2006.

DP

Datum odbrane:

DO

Članovi komisije:

KO

Predsednik:

Član:

Član:

Član:

UNIVERSITY OF NOVI SAD
FACULTY OF SCIENCES AND MATHEMATICS
KEY WORDS DOCUMENTATION

Accession number:

ANO

Identification number:

ino

Document type: Monograph type

DT

Type of record: Printed text

TR

Contents Code: Master thesis

CC

Author: Ivana Manić

AU

Mentor: Prof. Dr. Nataša Krejić

MN

Title: Mathematical Models for Estimation of Operational Risk and Risk Management

XI

Language of text: English

LT

Language of abstract:

LA

Country of publication: Serbia

CP

Locality of publication: Vojvodina

LP

Publication year: 2007

PY

Publisher: Author's reprint

PU

Publ. place: Novi Sad, Faculty of Science and Mathematics, Dositeja Obradovica 4

PP

Physical description:(5/111/32/39/11/21/0)
(chapters/pages/literature/tables/pictures/graphics/appendices)

PD

Scientific field: Mathematics

SF

Scientific discipline: Applied Mathematics

SD

Key words: Mathematical model in economy, risk management, operational risk, Value at Risk, Loss Distribution Approach, EM algorithm, EVT theory, Basel II

UC:

Holding data: The library of the Department of Mathematics and Computer Science, Novi Sad

HD

Note:

N

Abstract: This thesis is devoted to the mathematical models for estimation of operational risk in financial institutions. The estimation of operational risk is relatively new problem in the field of risk management and in this paper three basic approaches are considered and explained; Basic Indicator Approach, Standard Approach and Advanced Measurement Approach. The accent in the thesis is on the method Loss Distribution Approach which is based on different mathematical models such as Method Maximum Likelihood, fitting the data with heavy tailed distributions, algorithm Expectation-Maximization, Extreme Value Theory, Monte Carlo simulation, Statistical goodness-of-fit tests etc. The thesis have also empirical results for considered approaches and methods.

AB

Accepted by the Scientific Board: 20.03.2006.

Defended:

Thesis defend board:

President:

Member:

Member:

Member:

Biography

I was born on March, 30th 1980 in Čačak, Serbia, and since 1991 I live in Novi Sad, Serbia. In 1999 I enrolled at undergraduate program in mathematics at the Department of Mathematics and Informatics, Faculty of Science, University of Novi Sad.

In 2004 I finished the studies with average mark 9,12/10 and became Bachelor of Science in Mathematics - Financial Mathematics.

The same year I started the master studies at Faculty of Science, University of Novi Sad and obtained a scholarship from Ministry of Science, Republic of Serbia. In the year of 2005 I was at the Department of Mathematics "F. Enriques", University of Milan, Italy for one semester. My stay at University of Milan was financed by European Commission through Tempus Project CD JEP 17017-2002 Mathematics Curricula for Technological Development.

During the master studies I participated at "ECMI Mathematical Modelling Week"; in 2005 at University of Novi Sad and in 2006 at Technical University of Copenhagen, Denmark. Further, I have attended the course "Financial Mathematics" in Plovdiv, Bulgaria supported by DAAD. In September 2006 I have presented my work in the field of operational risk at the XVI Conference on Applied Mathematics in Kragujevac, Serbia. In 2007 I have participated in project Modelling and Forcasting Stock Process Behaviour - Analysis of High Frequency Data at the Department of Mathematics and Informatics, University of Novi Sad. The project was conducting for Dresdener Kleinwort Securities Limited London. In July 2007 I have been at Credit Risk Training in VUB bank Bratislava, Slovakia organized and financed by Panonska banka ad Novi Sad.

During 2007 I worked in "M&V Investments" brokerage house in Belgrade, Serbia as a financial analyst. Since December 2007 I work in UniCredit Bank, Belgrade as risk controller.

Novi Sad, 22.11.2007. Ivana Manić