Fraud detection using operational risk modelling with incomplete data.

by Muzerengwa Kudzai Calvin

February 14, 2019

Submitted in partial fulfillment of the requirements for the degree Master of Science (Actuarial Science) in the Faculty of Natural and Agricultural Sciences

University of Pretoria

Pretoria

© University of Pretoria

Declaration

I, Kudzai Calvin Muzerengwa declare that the dissertation, which I hereby submit for the degree of Master of Science (Actuarial Science) at the University of Pretoria, is my own work and has not been previously been submitted by me for a degree at this or any other tertiary institution.

Signature:...............................

Date:....................................

Supervisors - Professor Kijko and Dr Conrad Beyers

Department - Actuarial Science

Degree - Master of Science(Actuarial Science)

© University of Pretoria

Acknowledgements

Firstly, I would like to thank God for providing me the opportunity to pursue my studies up to this level. Secondly, I would like to thank MMI for providing me with the funds to do my studies. My supervisors, Dr F.J.C. Beyers and Prof. A. Kijko for guiding me and providing feedback on the progress I was making throughout my studies. Lastly, I would like to thank my family and friends for giving me emotional support that helped me throughout my studies.

© University of Pretoria

Abstract

Systems and processes may fail and employees can engage in fraudulent activities that can go unnoticed for a very long time and the resulting losses can be very high and catastrophic to an institution. Setting a minimum threshold or a level of completeness will not guarantee that all losses above this point will be reported. In order to model operational risk data, a method that does not depend on the level of completeness is suggested. This can be done by introducing a detection probability that is combined with the underlying loss distribution to give a 3-parameter gamma distribution and fitted to a simulated dataset. It is found that the methodology is able to accurately estimate parameters when the data is incomplete.

Keywords – Level of Completion, Loss Data Analysis, Operational risk, Gutenberg-Richter *b*-value,

© University of Pretoria

List of Abbreviations

- BI - Basic Indicator method
- BSCR - Basic Solvency Capital
- EU - European Union
- EVT - Extreme Value Theory
- LDA - Loss Distribution Approach
- MCR - Minimum Capital Requirement
- MME - Method of Moments Estimation
- MLE - Maximum Likelihood Estimation
- MBASS - Median Based Analysis of Slope Segment
- ORSA - Own Risk and Solvency Assessment
- POT - Peak-Over-Threshold
- SA - Standardised Approach
- SCR - Solvency Capital Requirement

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Contents

© University of Pretoria

© University of Pretoria

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

List of Figures

© University of Pretoria

© University of Pretoria

© University of Pretoria

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

List of Tables

© University of Pretoria

Chapter 1

Introduction

1.1 Introduction

There has been a number of financial institutions collapsing or incurring huge losses due to poor operational risk management practices. The collapse of Barings Bank in 1995 was due to Nick Leeson, a derivatives trader, who executed unauthorized trade deals over a period of four years. He speculated on the Nikkei 225 which became unstable following the Kobe earthquake in 1995 and this resulted in losses amounting to £926 million (R16.96 billion) [1]. Bernard L. Madoff in December of 2008, was arrested on charges of securities fraud. He ran a Ponzi scheme where he got funding from investors and deposited the funds into his bank account. He would go on to forge return statements for investors. This scheme had thousands of victims and the total losses were US$57 billion (R812.82 billion)[2]. Jerome Kerviel a French trader working for Societe Generale, in 2007, executed unauthorized trades which resulted in a total loss of €4.9 billion (R81.28 billion) [3]. Other operational risk losses include Merrill Lynch in 2013 (discrimination), Lehman Brothers in 2008 (poor advisory activities, improper business practice), NASDAQ in 2013(business interruption) to name but a few. Lack of proper operational risk management systems has been considered to be part

© University of Pretoria

of the cause to the 2008 financial economic crisis [4] and [5]. Unauthorized trading, internal and external fraud were part of the top 10 operational risk events for the years 2016 to 2017 [6].

Financial services regulators, since the Global Financial Crisis, have been putting measures in place to prevent another crisis from happening and mitigate the adverse impact if one were to occur. The Basel Committee on Banking Supervision, a committee of banking supervision authorities around the world, implemented the Basel Accord, which is a set of rules and principles which member states adopt when setting their own local banking regulations. Numerous non-banking sector regulators have adopted some of the principles set in the Basel accord, mainly Basel II, in their local regulatory frameworks. In Europe, there is the Solvency II framework which is a set of guidelines for European insurers to follow when calculating capital requirements.

The then Financial Service Board of South Africa, the regulator for the non-banking sector, decided to adopt some of the principles from Solvency II to develop the Solvency Assessment Management (SAM). SAM regulation came into effect on 01 July 2018. The main objective of this regulation is to ensure that insurers have enough capital to cushion them against the risk of insolvency by setting a minimum capital requirement for risks such as liquidity, credit, operational risk and others [7], [8] and [9].

Operational risk is less clearly defined compared to other risks faced by financial institutions such as credit and market risk. Up until recent years, operational risk was considered to be the risk class that contains 'other' risks. Now the most agreed upon definition of operational risk is the risk that an institution faces due to people, busi-

© University of Pretoria

ness, or system failure [10]. Examples of operational risks include internal fraud such as theft, external events, such as security breaches, damage to assets, business interruptions, poor employment practices and workplace safety [11]. There may be varying exposure to each subclass for different institutions and the degree of interrelationship among these risk sub-classes is not the same for all institutions.

The frequency and severity of losses an institution incurs due to an operational risk event give a presentation of the institution's risk exposure. By analyzing historical loss data, estimates can be made on how much buffer capital an institution needs to allocate against potential losses. The challenge is to find a statistical distribution that best models loss frequency and loss severity taking into account for losses that go undetected. Depending on an organisation's fraud detection systems, losses will go unreported or records can be falsified to cover up the losses. This will result in loss datasets that are incomplete. The level of result inaccuracy can be further exacerbated by the implementation of a minimum loss collection threshold which results in more losses being excluded from the dataset.

An alternative method would be to use the full dataset for capital modeling. At small loss amounts, the dataset is considered to be incomplete and this is further explained in Section 1.3 below. A probability of detection function is introduced to allow the minimum collection threshold be flexible to allow for some losses smaller than the minimum collection threshold to be able to be recorded and added to the dataset. Also the detection function will take into account some missing losses above the threshold that may not have been reported. The resulting distribution will give a parameter that describes the underlying loss distribution. There has been a significant move towards formula based methods under the new Basel regulation and so analysis of frequency-severity of losses is still useful to companies, specifically for internal risk management

© University of Pretoria

purposes.

1.2 Organization of dissertation

This dissertation is split into six chapters. Chapter 1 will cover the introduction, background to operational risk modelling, aim and motivation for introducing a model that uses a polynomial detection function for losses instead of assuming that all losses are reported and included in the dataset. In Chapter 2 a review of relevant literature of financial fraud cases and fraud detection is discussed. Financial industry regulation and solvency capital requirements are discussed in Chapters 3 and 4. The proposed loss amount distribution is explained in Chapter 5. Chapter 6 covers the modelling results from the simulation exercise done and the results are discussed in Chapter 7. Chapter 8 holds the conclusion of the dissertation.

1.3 Problem of incomplete datasets for modelling

When analyzing the distribution of a random variable, it is useful to consider the three main "parts" of the distribution which are: the lower tail, the body and the upper tail, as shown in Fig.1.1 [11]. Operational risk capital requirements are mainly set against the risk of unexpected losses and these losses tend to be of a high severity and low frequency nature. Large unexpected losses can ruin a company and the effects of such an event can ripple throughout the industry posing systemic risk. This makes institutions focus more on understanding the behaviour of the upper tail of their loss distribution.

Losses that are of low severity high frequency usually do not get special attention as they are believed to be preventable and exhibit expected behaviour [11]. As the sever-

© University of Pretoria

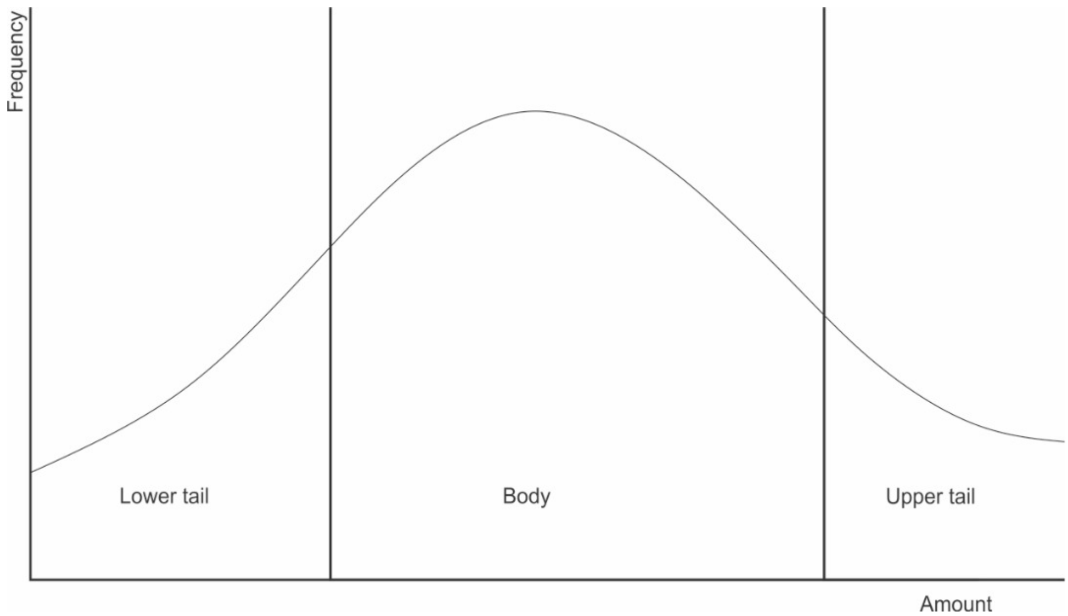


Figure 1.1: Three main "parts" of a loss distribution: lower tail, body and upper tail.

ity of a loss increases, the likelihood of the loss being noticed and recorded increases. At smaller loss amounts, not all losses are recorded, since some go unnoticed either because they are intentionally hidden or they are considered not to be significant enough to include in the dataset. The relatively small losses may be disregarded in some large institutions as they may be considered not material to influence the solvency level of the institution. Institutions will often neglect the lower tail when sometimes the lower tail may be highly significant. Fraud is an example of operational risk that is difficult to detect because perpetrators tend to hide their tracks and so the losses may not be very detectable below certain thresholds. There are two main thresholds, minimum collection threshold is a set minimum loss amount that is recorded and added to a datset [12]. Then there is the level of completeness which is the minimum loss amount where all losses above this point are detected [13]. Some losses can be much higher than the minimum collection threshold which will be but go undetected and hence unreported. A company's reporting standards may allow for losses at lower amounts to go unreported. Fraud in financial institutions may result in loss datasets that are incomplete.

© University of Pretoria

Under Basel II, banks are advised to set €10 000 as a minimum data collection threshold, which means only losses above this point are recorded [12]. This converts to about R158,000 for a South African institution. A dataset is considered to be complete if for each loss amount, all the losses that occur are observed and recorded. Using a complete dataset to model the loss severity and frequency would be ideal as this will give the true distribution of losses.

In practice, it is observed that at some loss amounts, some losses are not recorded because they are hidden or they are too small to be noticed. Above the level of completeness, all losses of amount greater or equal to this point are observed and recorded. The dataset is then considered to be incomplete since some losses that occurred will be missing at some severity levels. An operational risk dataset is then assumed to be complete for all losses above the minimum collection threshold which will be referred to as the level of completeness. To set the distribution of the losses, an unconditional distribution is fitted to the data. This assumes that the losses recorded are the only losses that have occurred. This approach results in a biased distribution because the losses below the threshold are not included in the dataset, as shown in Fig.1.2.

A more appropriate approach is to fit a conditional distribution to the data [14]. This takes into account the fact that the data recorded is above the collection threshold and the resulting distribution will only be for the losses above this threshold point. As in Fig 1.2, the losses below the level of completeness or threshold are not included in the modelling of operational risk. The dataset will be truncated and information of losses below this point is not taken into account. Without truncation, the resulting distribution will be distorted because the dataset used is incomplete and not a good representation of the actual losses that are incurred. Operational risk data collection exercises have only started in recent years and so the datasets are typically relatively

© University of Pretoria

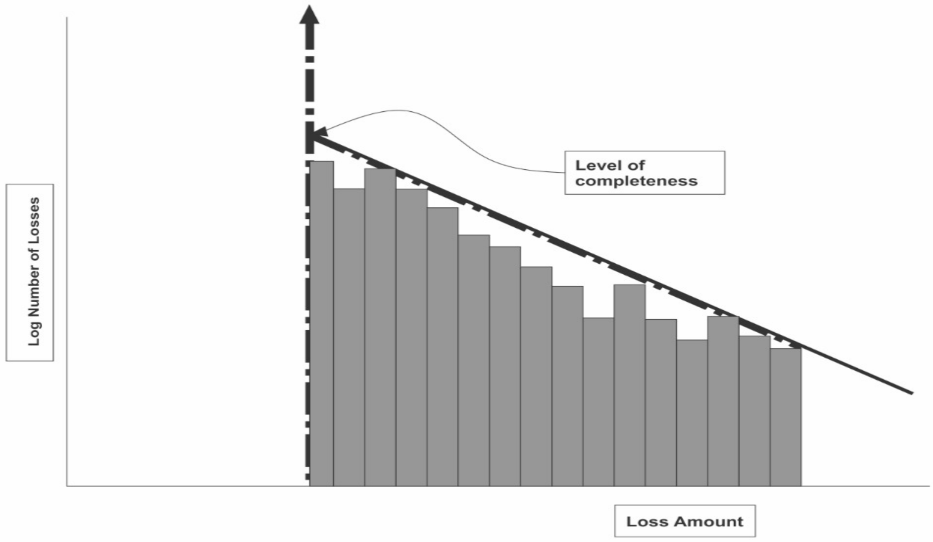


Figure 1.2: Truncated distribution that does not take into account the losses below the level of completeness [11].

small. Statistical modelling depends mainly on fitting a large dataset to get a distribution that closely models the real world scenario. Since the datasets are small to begin with, further truncating them significantly reduces the credibility of the resulting distribution.

Another consideration for having a threshold is that institutions have different structures and the risks they face are different, and so setting a single threshold may be unsuitable. A general insurer, a life insurer and a commercial bank will have different operational loss profiles and so setting a single minimum threshold for all three institutions may result in over- or undercharging the capital requirements [15]. Also a different threshold may apply for different business classes and different operational risk sub-classes. Individual institutions are allowed to set their own internal thresholds by changing the minimum threshold level and investigating the effects on the level of capital requirements. The regulator will require institutions to disclose the reasons behind their model assumptions [12]. The data collection threshold or a modelling threshold, has a significant impact on the modelling results. An institution may

© University of Pretoria

experience numerous operational risk losses of small amounts that do not reach the threshold level. The individual losses may not make an impact on the solvency of the institution, but the aggregate may have a significant impact. These losses will need to be recorded and taken into account when defining the distribution of operational risk losses. The threshold can be set too low such that the added cost of data collection outweighs the benefits of incorporating the data in the modelling exercise. Threshold levels can result in parameter estimates that are less robust and have higher bias [16].

1.4 Significance of loss distribution thresholds

The Loss Distribution Approach (LDA), an approach of analysing the risk level based on historic losses faced by an institution, relies on the historical loss data on operational risk and so data collection and reporting systems need to be put in place. Instead of truncating datasets, it is conservative to utilize all the data loss that was observed in modelling the distribution of losses. Understanding the full distribution and being able to model losses accurately is important, even at small losses where data is known to be missing. An incomplete dataset can give information that is useful compared to a truncated dataset. Inference can be carried out to see how much losses go unreported and why it it is the case. The challenge of picking the truncation point can result in institutions having lower or higher reserves than needed as this affects the amount of data used in modelling exercises. Modelling operational risk using an LDA approach that accounts for the full distribution of the losses faced by an institution, helps set accurate levels of capital requirements for operational risk and internal risk management that can give an institution a competitive edge over its peers. Losses that fall below the level of completeness can significantly change the estimated aggregate loss. Total losses will be lower if the dataset is assumed to be complete when it is not and when included in modelling data resulting in parameters with less bias and better fit.

© University of Pretoria

A threshold is defined as a magnitude or intensity that must be exceeded to trigger a specific action. In this case, intensity or magnitude refers to loss amounts and loss amounts that are greater than the threshold will be observed, recorded and included in the modelling exercise. The threshold has to be set correctly so that only enough relevant data is utilized. The threshold level applied to a dataset has a significant impact on the resulting distribution. Thresholds can be put in place because of the cost of data collection. If the cost is too high, it may be in the best interests of the company to leave out some losses as the cost will not be worth the benefit. An institution with poor reporting practices may render data at small losses less credible if losses are not properly reported. The effects of having a threshold can result in inadequate capital reserves.

To help explain this, an example using 1000 random variates simulated data from a lognormal distribution with parameters $\mu = 3,5$ and $\sigma = 0.5$, i.e. mean value of about 37.52. A truncated lognormal distribution is fitted to a range of threshold values. The results are shown in Fig. 3. As the threshold level increases, the parameter volatility increases together with the mean of the distribution. This is shown by the data in Table 1.1. Truncating data results in higher averages as shown in Table 1.1.

To assess the impact a threshold can have on Value at risk (VaR), a widely used risk measure, the simulated datasets are used to create new aggregate distributions. The datasets will represent losses and loss frequencies derived from a Poisson distribution with mean of 20 and the results are displayed in Fig. 1.4. The results assume that the distribution of loss frequencies does not change as the threshold changes.

From this, we see that as the threshold level increases, VaR increases as well. By hav-

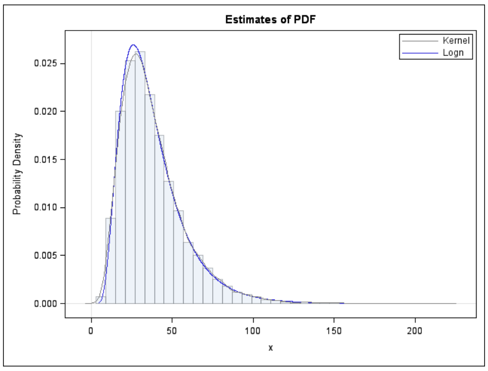
© University of Pretoria

Threshold	μ		σ		MEAN
	estimate	st dv	estimate	st dv	
N/A	3.51096	0.00502	0.50165	0.00355	37.9696
20	3.531	0.00887	0.48766	0.00629	38.4711
30	3.52842	0.02305	0.48924	0.0112	38.4016
40	3.56919	0.04779	0.47519	0.018	39.7295
50	3.67149	0.07802	0.4468	0.02571	43.4367
60	3.73307	0.13134	0.43156	0.03798	45.8874

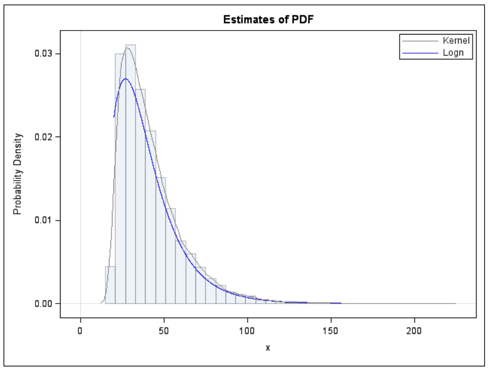
Table 1.1: Parameter estimates and their respective standard deviation from a sample of 1000 from a lognormal distribution with $\mu = 3.5$ *and* $\sigma = 0.5$

ing a threshold of 60, VaR increases by almost 70% compared to a scenario with no threshold. This may be prudent as companies will have to set high reserves, but in some cases the reserves may be too high, locking capital in. There is a need to better estimate the loss distributions without setting the threshold.

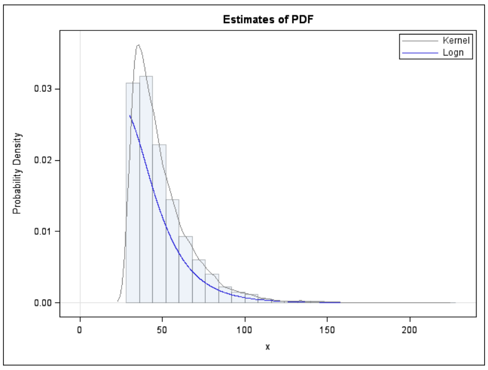
© University of Pretoria



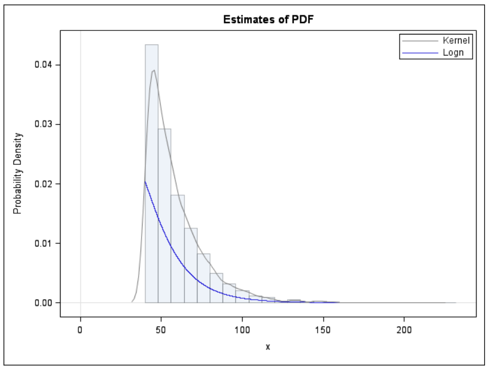
(a) True distribution with no truncation



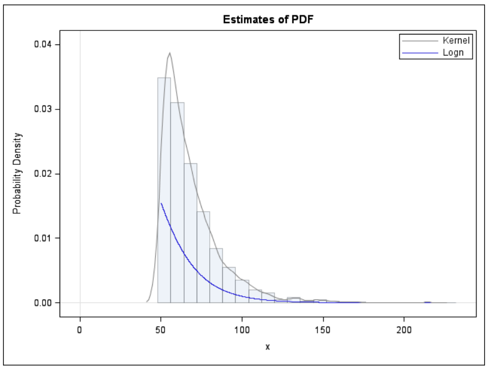
(b) Threshold of 20



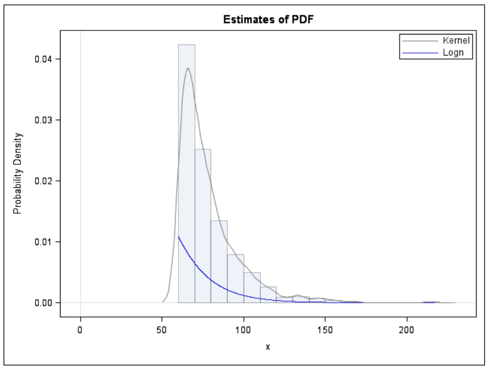
(c) Threshold of 30



(d) Threshold of 40



(e) Threshold of 50



(f) Threshold of 60

Figure 1.3: Plots showing the resulting distributions under different levels of truncation

© University of Pretoria

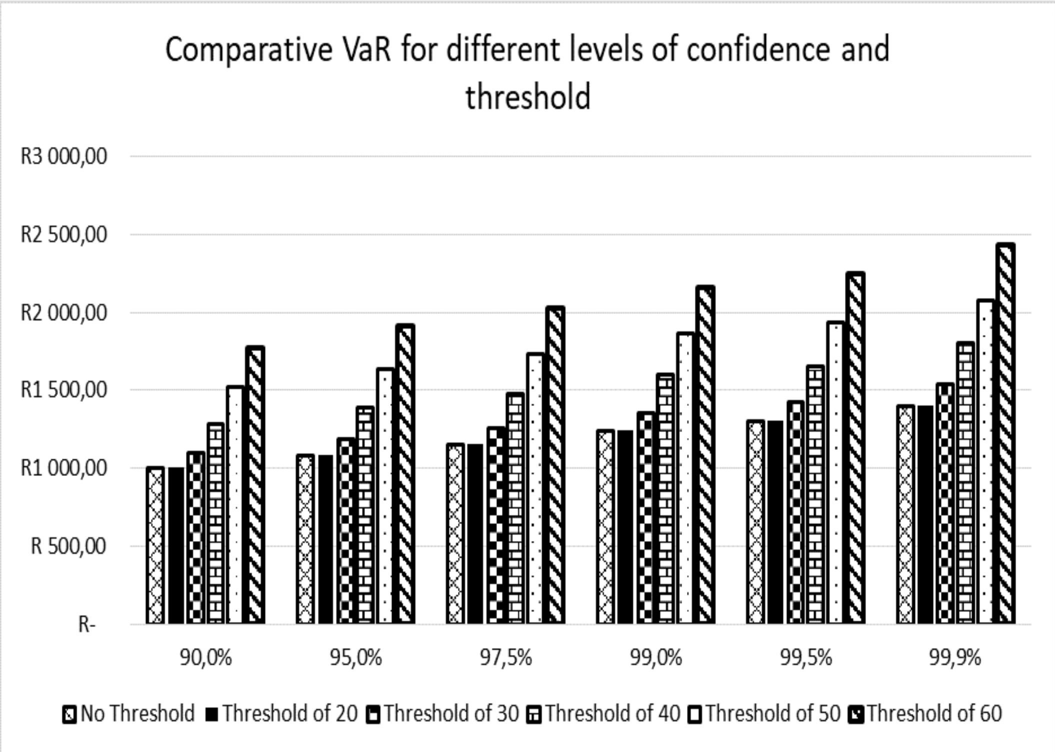


Figure 1.4: The comparative Value at Risk and effects of different thresholds.

© University of Pretoria

1.5 Research question

The level of completeness is a significant factor to consider when modelling losses but it is not clear on how to accurately estimate it. Under Basel II, the minimum collection threshold is estimated on a purely subjective basis and banks can set their own thresholds. There is a need to introduce a way that takes into account the features of the available data for the modelling exercise. A similar problem is found in seismology and a solution has been proposed to model earthquake frequency-magnitude distribution without specifying a threshold. In this study, a detection probability is used instead of setting an estimate of the level of completeness and the model is applied to operational risk data.

1.6 Overview of methodology

A quantitative approach is adopted for this research. In order to set adequate capital reserves, there is a need to accurately model the loss patterns. And so, a statistical approach is adopted to set the distribution of historic losses.

This is inter-disciplinary research with the methodology developed for seismic hazard modelling now being adopted and modified for operational risk modelling. This approach was chosen because operational risk losses and earthquake magnitude have similar distribution structures such as "high frequency with low impact" and "low frequency with high impact" and there are hidden events. There are earthquakes that go unnoticed just as there are fraudulent activities that result in losses but are kept hidden. A statistical distribution that does not make assumptions about a minimum data threshold [17] is developed and fitted to a simulated dataset and historic loss data on operational risk.

© University of Pretoria

1.7 Rationale and significance of study

Earthquake data collection started as early as 1900 while operational risk data collection is still in its infancy. For both, the datasets have a specific severity point where data of lower severity are incomplete. For earthquake datasets, the data is missing due to the limitations of the instruments or because some small earthquakes are not naturally occurring. For operational risk, small risk events are not recorded because they may go unnoticed. For both cases, the datasets are truncated at the minimum observable earthquake magnitude, m_{min} and collection threshold point, minimum loss amount that can be reported, for operational risk.

Thresholding, process of setting a minimum threshold, is done in both earthquake and operational risk modelling. There is mainly hard thresholding in earthquake magnitude modelling with data above m_{min} included in datasets used for modelling. In operational risk, there is believed to exist a point at which the distribution of the body differs from the upper tail. The m_{min} and collection threshold differ from the level of completeness as the level of completeness is the point where all values from this point onward are observed but with m_{min} and collection threshold some values bigger than these go unnoticed.

Operational risk modelling depends heavily on historical data on losses and so there is a need to analyze all the data available. Instead of focusing on one end of the distribution, the proposed method also gives a clear distribution of the small losses as well and so losses that may be considered too small when they are actually significant will be picked up. This method focuses on assessing the impact and frequency of all losses, regardless of the size. Even though regulatory operational risk requirements will be calculated on formula based methods, it is still necessary to investigate the loss

© University of Pretoria

frequency and loss severity for internal management purposes.

© University of Pretoria

Chapter 2

Literature Review: Financial fraud

2.1 Introduction

The global financial service industry has greatly increased in terms of level of sophistication due to the increased use of financial derivatives [18]. Financial products can now be transformed from one form to the other, for example through securitization. This has resulted in numerous new and complicated financial instruments introduced on to the markets and the connection among financial institutions has become stronger. If one institution fails, the ripple effect it has on the whole system could be high and very unpredictable. The use of the internet to execute trade deals and collect and store data poses new risks and opportunities to the industry.

Financial services regulation has improved in the last 30 years. In order to make sure that financial service providers are able to meet their obligations to customers, regulators prescribe these institutions to set aside capital to protect against credit risk, market risk, and recently operational risk. The regulators' goal is to ensure financial stability in the financial sector, fair markets, reduced financial fraud and to protect customers. A number of methods to calculate capital requirements have been proposed and this

© University of Pretoria

chapter covers the background to these methods, with special attention to capital for operational risk and methods that use level of completeness.

Even though regulation has been the biggest driver of risk management frameworks within institutions, there are other influences that are becoming more important. Institutions calculate regulatory as well as economic capital. Regulatory capital is calculated under stringent rules and regulations set by the regulator. Economic capital is defined as the the surplus assets a company requires to remain solvent for a specified time period at a specific risk level [19]. Unlike regulatory company, economic capital calculations are based on the individual company itself and its risk profile. Loss distribution assessment is an important exercise for economic capital calculation because institutions need to be aware of the true losses they made, and they expect to make in the future. A rigorous loss assessment can provide a competitive edge and reduce the risk of reputation damage.

In this chapter, The literature on fraud as one of the operational risk is given which focuses on the notable fraud cases and fraud detection techniques. An overview of common method used to model operational risk is provided together with the regulatory capital requirement calculation approaches.

2.2 Operational risk example: Fraud

Financial fraud is defined as "a deliberate act that is contrary to rules, law or policy with intent to make an unauthorized financial gain" [20] and [21]. Fraud is a big concern for customers, financial institutions and regulators of financial industries. Whether it is external or internal fraud, it can lead to ruinous results and potentially have a knock on effect that lead to further losses within the industry. This is in deed

© University of Pretoria

one of the most important operational risks.

Financial fraud ranges from small scale credit card fraud to big cases like market manipulation. Financial fraud is subdivided into 4 main groups, namely bank fraud, insurance fraud, securities fraud and others [22]. Bank fraud is the act of knowingly executing a scheme to defraud a financial institution. This covers credit card fraud and mortgage fraud which will fall under external fraudulent activities where third parties try to gain funds from the bank or other customers. Money laundering is of serious concern to regulators and security authorities. Money laundering is when money is deposited into a bank account under the pretense that it was legally obtained when in fact it was earned through illegal activities. These activities are mainly done by the employees or business owners, and so they form part of the internal fraudulent activities [23].

Insurance fraud is when consumers of insurance products either provide false information at application, rating, billing or claiming stages of the insurance process or policyholders try to make fraudulent claims. This mainly forms part of the external fraudulent activities that insurers record.

Securities and commodities fraud is the class that have some of the biggest fraud cases that can result in very large losses. This is a class of fraud that can be committed by the misrepresentation, falsification or manipulation of information in order to gain from financial investment decisions [24] and [25]. This involves market manipulation, Ponzi scheme, pyramid scheme, hedge fund fraud, embezzlement among others. The losses that can occur under this group can range in the billions of dollars and it can take time to detect the fraudulent activities going on.

© University of Pretoria

The last group is what is considered "Other" types of fraud. This covers mainly corporate fraudulent activities such as financial statement falsification and insider trading. Also, it covers fraud carried out via mass-media communications such as internet and telephone.

2.2.1 Notable fraud cases

Many notable cases of operational risk include the collapse of the Barings Bank after an employee performed unsanctioned trades, the huge loss suffered by Sumitomo after an employee did rogue trading for a period of 10 years without getting noticed and Bernie Madoff who defrauded his clients of hundreds of millions in a large scale Ponzi scheme. In more recent times, there was the collapse of the Woodridge Group, a real estate firm that defrauded its investors of US$1billion, A Japanese bank, Shoko Chukin, where employees across several branches falsified documents to increase lending.

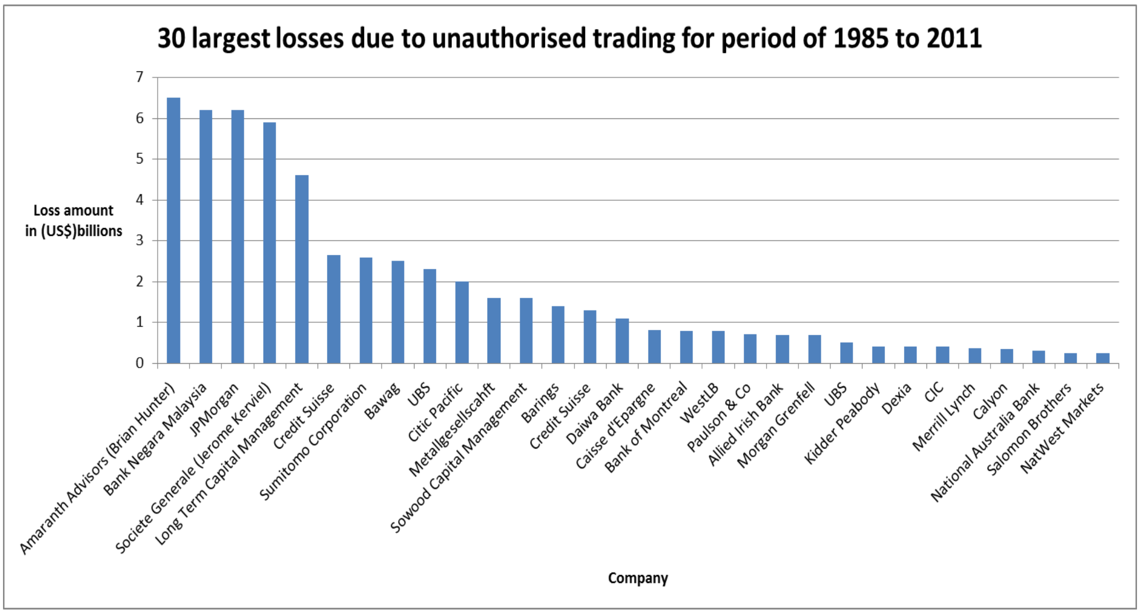


Figure 2.1: 30 largest loss amounts on the fraud cases collected by Willis RE.

Domestically, the most prominent financial fraud case involves the Gupta family busi-

© University of Pretoria

ness which is linked to other institutions that include the government, McKinsey and KPMG [26]. KPMG plunged into more trouble after lead auditor responsible for VBS Mutual bank failed to fully state the extent to which he was involved with the bank [27]. VBS bank was later put under curatorship by the Reserve Bank. Steinhoff International is another company that has been involved in financial fraud. Senior executive officers have been alleged to have acquired companies off the book in order to set them up to hide losses [28] and [29].

2.3 LDA on operational risk modelling

Obtaining a good understanding of the losses an institution faces is one of the most important operational risk management goals. In order to take into account the risk profile each institution faces, the Basel Committee updated the Basel Accords in 2004. Under these updates, there was an option to use loss distributions to calculate capital. These methods can be used not only to calculate regulatory capital but for internal management as well.

The impact of an operational risk event is felt as a financial loss on an institution's accounts. Under the LDA approach, the main goal is to analyze the distribution of individual loss amounts and loss frequency so that the aggregate distribution of losses is used to estimate the capital charge. Institutions are given the opportunity to construct their own internal models and use internal data in calculating the capital requirements. The internal data that is used in the modelling process can be supplemented by external data to improve reliability because the internal data volume may be too low. This method is considered to be rigorous because losses are considered to be a good risk indicator compared to other indicators such as gross income used in other approaches.

© University of Pretoria

Because institutions construct internal models, regulators set prerequisites that institutions must meet before they can adopt this approach. To calculate capital charge, institutions calculate the VaR at a specific level of confidence. According to Basel II, a bank need to hold capital that provides a 99.9% level of confidence that the bank will not go insolvent within one year [30] and under Solvency II, capital needs to provide a 99.5% level of confidence [31]. Frequency-severity distribution analysis is one of the most important methods adopted under the Advanced Measurement Approach (AMA) to model operational risk.

2.4 Loss frequency distributions

Operational risk losses can be grouped loosely into two main categories. The first group includes expected losses that occur more frequently. The second group includes losses that rarely occur and less predictably. The frequency at which the losses occur is random and so in order to model operational risk there is a need to model the loss frequency [32], [11] and [33]. Discrete distributions are used to model loss frequency and the assumption that loss frequencies are independent is usually made. The following distributions are the most commonly fitted frequency distribution.

Poisson and negative binomial distributions

The Poisson distribution assumes that the number of losses is unbounded. In a specified time period there is an average of events that will occur and this is called the intensity rate. For a standard Poisson distribution, the intensity rate remains constant. It is more realistic to assume that as time changes, the intensity rate changes as well and in this case it is called a non-homogeneous Poisson process. This is the most commonly used model because it does not limit the maximum number of losses. The sum

© University of Pretoria

of Poisson random variables gives a Poisson random variable. This is useful when aggregating the total number of losses over different risk classes and business units [11].

The negative binomial distribution is a special case of the Poisson distribution. The intensity rate of the Poisson distribution is no longer constant but a stochastic random variable following a Gamma distribution. This distribution gives greater flexibility in modelling number of losses occurring in a specified time period. Unlike the Poisson, the intensity rate can assume a range of values instead of remaining constant. The geometric distribution is a special case of negative binomial where r the number of failures is set equal to one and this model can also be used to model loss frequency [34].

Cruz [35] fitted the Poisson and negative binomial to a large bank fraud dataset. The data set had 3,338 entries collected from 1992 to 1996 modelled as number of events per day.

Distribution	Parameters
Poisson	$\lambda = 2.379$
Negative Binomial	$r = 3.52$ and $\beta = 0.67737$

Table 2.1: Parameters of fitted distributions.

Cruz's results are shown in Table 2.1 and Fig. 2.2 and concluded that the Poisson was a slightly better fit and had the advantage of being computationally simpler to calculate the aggregate distribution.

Moscadelli [36] fitted the Poisson and negative binomial distributions to data collected by the Risk Management Group of the Basel Committee in 2002 Loss Data Collection Exercise from 89 banks. The dataset had 45,569 observations which were split into 8 business lines. Negative binomial had a better fit compared to Poisson despite data

© University of Pretoria

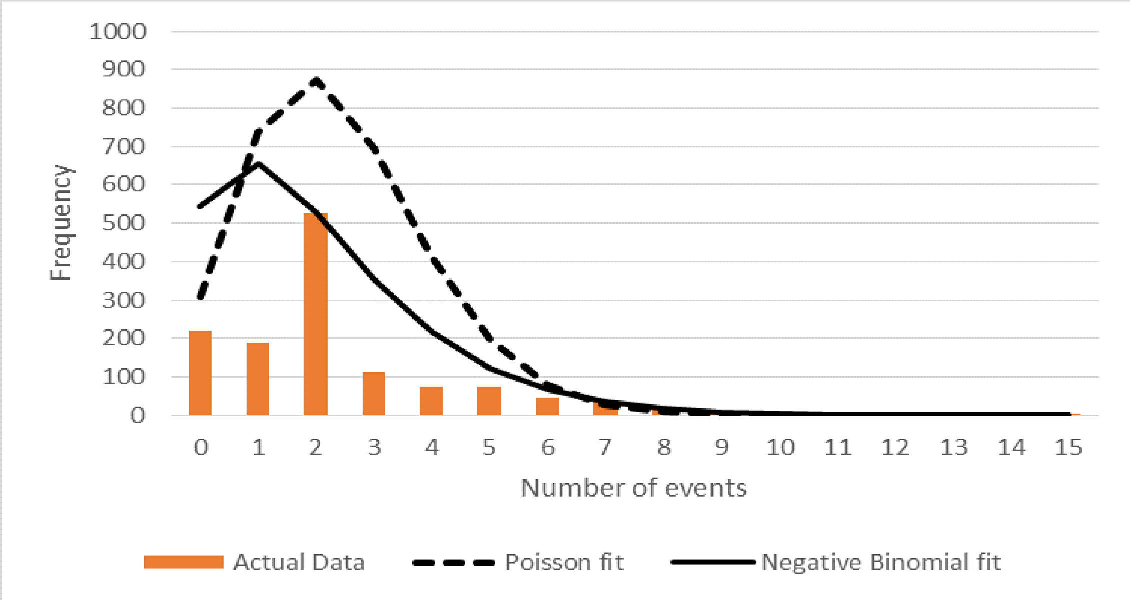


Figure 2.2: The actual data and fitted distributions.

being skewed to the right. De Fontnouvelle and others [37] pointed out that the Poisson distribution may not be the best distribution since it is not dependent on time that lapsed since the last loss.

2.5 Loss amount distributions

Operational risk loss amounts are uncertain and the information on them is difficult to acquire in full because of difficulties in identifying all data and recording it correctly. Data can be modelled using a non-parametric approach in which empirical data is used or a parametric approach can be adopted where an analytical distribution is fitted to the data.

Operational risk data is highly skewed to the right and so in most cases, the log-data is fitted to distributions. In some cases, data is split into two sets, one for smaller losses and the other for large losses. They are then modelled separately.

© University of Pretoria

De Fontnouvelle and others modelled severity of the 2002 Operational Risk Loss Data Collection (LDCE), the same dataset modelled by Cruz [35] and Moscadelli [36]. In their study, they focused on six banks that contributed data to this dataset. The data was split according to business lines and event type based on the Basel classification system. The following distributions were fitted: lognormal, Burr, exponential, Pareto, GDP, loglogistic, Weibull and gamma distributions [38]. Using a chi-squared goodness of fit, they concluded that the heavy tailed distributions fitted the data better. These include Burr, lognormal, Pareto and GDP while the medium-tail distributions did not fit the data well.

In 2004 Dutta and Perry modelled data collected from the 2004 LDCE, and they fit the exponential, Weibull, gamma, truncated lognormal, loglogistic and Generalised Pareto distribution (GPD) to data from six banks. In addition to these they fitted two other distributions which they called the generalised beta distribution of the second kind(GB2) and the g-and-h distribution [39]. The GB2 distribution is defined as

$$f_X(x) = \frac{|a|x^{ap-1}}{b^{ap}B(p,q)[1+(\frac{x}{b})^a]^{p+q}} I_{0,\infty}(x), \tag{2.1}$$

where a, b, p and q are the four parameters of the distribution, $B(p,q)$ is the beta function and $I_{0,\infty}(x)$ is an indicator function. The parameters a, p and q are the shape parameters and b is the scale parameter [40].

The g-and-h distribution is a transformation of the standard normal distribution and it is defined as

$$f_{g,h}(Z) = (e^{gZ}-1)\frac{e^{\frac{hZ2}{2}}}{g}. \tag{2.2}$$

The parameter g is a real constant and it controls the level of skewness and $h \geq 0$ control level of kurtosis and elongation of the distribution [41].

© University of Pretoria

In their finding, the exponential, Weibull and gamma were poor fits for all banks. Loglogistic and GPD did not fit well for some banks while it fitted well for the other banks. The g-and-h distribution fitted the data well and had a reasonable capital charge as well.

2.6 Compound distribution for operational risk losses

Compound distribution take into account the random nature of the number of a random variable. This makes them ideal to insurance claims data [42] and operational risk data [11]. Let $X_1, X_2, X_3, ..., X_{N(t)}$ be a set of independent and identically distributed random variables where X_i and $N(t)$ are independent for all i and all t. $S(t)$ is a compound random variable defined as,

$$\begin{aligned} S(t) =& X_1 + X_2 + X_3 + ... + X_{N(t)} \\ =& \sum_{i=1}^{N(t)} Xi. \end{aligned} \tag{2.3}$$

The resulting distribution is a stochastic process and is used to represent the aggregate operational risk loss amount. The distributions discussed in Section 2.4 are used to model $N(t)$, the loss frequency and the loss amount of the i^{th} loss will be modelled by distributions discussed in Section 2.5. The most common distributions to be applied are the Poisson Process and Pareto distributions.

The mean and variance for this compound distribution are defined as

$$\begin{aligned} E\left(S(t)\right) =& E\left(Xi\right) E\left(N(t)\right) \\ Var\left(S(t)\right) =& E\left(N(t)\right) Var\left(Xi\right) + Var\left(N(t)\right) \left[E\left(Xi\right)\right]^2 \end{aligned} \tag{2.4}$$

This distribution is used to calculate capital charge [43]. This distribution is then used to calculate the Value at Risk. The biggest challenge with this method is that it does not

© University of Pretoria

account for the correlations between loss amount and loss frequency and correlations between different classes.

2.7 EVT in operational risk modelling

Operational risk modelling focuses mainly on modelling unexpected few large losses that can result in complete ruin of an institution. To model the distribution of these extreme events, Extreme Value Theory (EVT) is implemented. This theory was developed by Leonard Tippett and Ronald Fisher [44] and has been applied not only to operational risk modelling but to science, engineering, insurance [45], risk management [46] and finance [47] and [48]. There are two main ways of implementing EVT.

2.7.1 Block maxima

The first method is called the Block maxima method. In this method, data is divided into discrete blocks and the maximum of each block is considered to be the only observed value. In the end there is a new data set of the maximum of each block and an extreme value distribution is fitted on to this new data set. Let X_n be the maximum of a block of size n and $F_n(x)$ be the distribution function of X_n. As n approaches infinity $F_n(x)$ approaches H defined as,

$$H_{\xi,\mu,\sigma}(x) = \begin{cases} e^{-(1+\xi(\frac{x-\mu}{\sigma}))^{-1/\xi}} & \xi \neq 0 \\ e^{-e^{-(\frac{x-\mu}{\sigma})}} & \xi = 0 \end{cases} \tag{2.5}$$

μ and σ are normalising constants and ξ is the shape parameter. If $\xi = 0$ then the distribution simplifies to Gumbel distributions that are short tailed such as exponential and normal distributions. If $\xi < 0$ then the distribution simplifies to a Weibull distributions that are heavy tailed such as Pareto distribution. And if $\xi > 0$ then the

© University of Pretoria

distribution simplifies to Frechet distribution [49].

2.7.2 Peak-over-threshold (POT)

The second method is the Peak-over-threshold (POT) method. Under this method, a threshold is set and only values that exceed this threshold will be considered to have been observed. Let X represent loss amount and u represent the threshold. The new random variable Y, defined as

$$Y = X - u \qquad X \geq u, \tag{2.6}$$

represent the excess over the threshold and these are the ones that will be modelled using EVT.

$$F_u(y) = P(X - u \leq y | X > u) \qquad 0 \leq y < x_0\,, \tag{2.7}$$

where $x_0 \leq \infty$ is the upper bound of the distribution. As the threshold is increased and approaches the upper bound, the distribution of excesses over the threshold converges to a Generalized Pareto distribution defined as

$$G_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \xi x/\beta)^{-1/\xi} & \xi \neq 0 \\ 1 - e^{(-x/\beta)} & \xi = 0 \end{cases} \tag{2.8}$$

This distribution is called the Generalised Pareto Distribution (GPD) because for different values of ξ the distribution simplifies into different distributions. For $\xi = 0$ $G_{\xi,\beta}(x)$ is a Gumbel distribution, data has a tail that express exponential behaviour, for $\xi > 0$, it is a Pareto distribution and data has a tail that shows polynomial behaviour and for $\xi < 0$ it is a Weibull distribution with endpoint defined as $\frac{\sigma_u}{|\xi|}$ [49]. The choice of the threshold is a constraint for this method.

© University of Pretoria

2.7.3 Issues with EVT

EVT is used to assess the extremely large values but it has its on pitfalls. First, operational risk datasets are still small and by implementing this method, more data points are lost [50]. It directs attention to the extremely large losses and does not focus on the smaller ones as well. These are just as significant to the financial health of an institution. There is generally low volume of extremely large loss data and so simulations are done to augment data volume used in modeling using EVT. The introduction of simulated data can lead to spurious results. Also, there are limitations to the selection of thresholds when implementing the POT method. With the Block Maxima method, there is need to decide on the block sizes. The threshold used may be inappropriate or may change with time [51].

2.8 Fraud detection

Internal auditing has been the long-standing method of detecting fraud in companies. Regulators both in the banking and non-banking sectors have strove to make sure that institutions are thoroughly audited by competent professional and have regular audits. Both Basel II and Solvency II have set out the aims, key people, scope and considerations of the internal audit practices. However, this is now proving to be inadequate since the remarkable expansion and sophistication of the financial industry [52] and [53].

2.8.1 Data mining techniques for fraud detection

As the financial industry becomes more sophisticated and complex, there has been a remarkable increase in the volume, velocity and variety of data being produced. This has led to the implementation of data mining techniques in fraud detection. Data

© University of Pretoria

mining is defined as the process of examining large datasets to discover new valuable information [54]. This process makes use of mathematics, statistics, artificial intelligence and machine learning to create an analytical tool that can assist auditors to detect financial irregularities during the audit process [21]. There are several techniques employed to try detect financial irregularities.

Classification is the process of classifying data under defined, discrete and un-ordered classes. The main aim is to classify data based on common patterns and traits exhibited. Classification is the most commonly implemented data mining application. This is achieved through the use of Neutral Networks, Bayesian classification, Decision tree inductions and Support Vector Machines (SVM) [55] and [21]. This method was used to detect factors that contributed to financial statement fraud in a group of 76 Greek manufacturing firms. Of these 76 firms, 38 were positively identified to be involved is some sort of financial statement fraud [56]. The results showed that Bayesian Belief Network, Neutral Network and Decision Tree models correctly classified the sample with accuracy of 90%, 80% and 74% respectively.

Clustering is the process of splitting data into clusters where constituents of each cluster are similar in terms of their features and characteristics. This differs from classification by it being an example of unsupervised learning. Unsupervised learning is when an algorithm is used to infer about a dataset without pre-specified response values. Under classification there are no predefined classes but instead one will be splitting a large dataset [57]. Glancy and Yadav [58] carried out research to assess whether fraud can be detected from documents submitted to the Securities and Exchange Commission (SEC). They extracted information from the 10-K financial statement of companies that were issued an Accounting and Auditing Enforcement Release (AAER) by the SEC, and they performed supervised expectation maximisation clustering and hierar-

© University of Pretoria

chical clustering. They managed to show that expectation maximisation was unstable but hierarchical clustering managed to identify likely fraud at 1% level of significance.

Regression is the process of deriving the relationship between a dependent variable and a set of other independent variables. This is also a common application used. Types of regression models used include both linear and non-linear models together with multivariate models [57]. Likelihood of a fraudulent activity happening will be taken as the dependent variable and for independent variables, they can range from accounting ratios, company size, technology, skill level of employees among others [59]. To the same group of companies modelled by Kirkos and others [56]. Spathis [52] fitted a regression model where he used accounting ratios, and he found that the following were significant towards the likelihood of false financial statements:

- net profit/ total assets
- working capital/ total assets
- gross profit/ total assets
- total debt/ total assets
- Z - score

where [60]

$$\begin{aligned} Z = {} & 1.2(working\ capital/\ total\ assets) \\ & + 1.4 * (retained\ earnings/\ total\ assets) \\ & + 3.3 * (earnins\ before\ interest\ and\ tax/\ total\ assets) \\ & + 0.06 * (market\ value\ equity/\ total\ debt) \\ & + 1 * (sales/\ total\ assets). \end{aligned} \tag{2.9}$$

Prediction is the process of estimating numeric and ordered continuous values based

© University of Pretoria

on patterns exhibited by data. This is closely related to regression where regression models are used to model likelihood of detecting fraud [53]. Another common prediction technique is the 'next to neighbour' techniques where the cluster or value of the next item is predicted based on similar items in database [61]. Outlier detection is the process of assessing how some data points greatly vary from the rest of the body of data.

Visualization is the process of presenting information gained from the process of data mining. By using diagrams, colour and other visual effects, complex ideas and information can be better understood [62]. The most commonly used technique is the self-organizing map (SOM) [63] and [64] which converts multidimensional data to a simpler 2 dimension. Simunic [65] used this method to visualize a cluster of stock charts. Huang and others [66] used this method to survey the market performance and combined it with a behaviour-driven analysis of trading networks to identify suspicious activity. Olszewski [67] used the SOM combined with threshold-type binary classification to detect fraud in telecommunications and credit card frauds. Argyriou and others [68] in their paper used a spiral system where events are plotted on the spiral based on the time at which they occur. Suspicious events are ones that are close to one another or close to the centre.

2.8.2 Limitations of data mining techniques in fraud detection

Data mining techniques can be used to aid risk managers through identifying high risk areas. These techniques can help identify losses hence making loss datasets more comprehensive and a step closer to completeness. There are limitations to these techniques, specifically privacy and co-operation, making them complicated to implement for fraud detection. And so there is need to set thresholds correctly. The larger the

© University of Pretoria

loss, the more likely it is observed however some large losses may be hidden and go unnoticed. Some of the major limitations to datamining techniques are listed below.

- **Short track record**: Data mining is a fairly new way of detecting fraud and so fewer people are familiar with the processes and benefits of utilising this in fraud detection. Data mining requires large volumes of data and current data sets on fraud and other operational risk events are still in small.

- **Practicality and applicability**: Different companies have different database systems. The type of data available may influence the data mining method implemented without regard to the risk management needs of the company. Also, some techniques such as Convolutional Neutral Networks are very complicated such that they cannot be implemented of available databases.

- **Cost** - data mining requires of computer power as well as trained operatives. This can be very costly to some companies such that they may decide that the benefits got will not be worth cost.

- **Scope**: Data mining techniques are not able to explain, predict or pick up all the fraudulent activities going on in a company. There will be need to implement other methods to supplement this [21].

- **Privacy**: Data mining requires lots of data about companies and customers. This can lead to privacy issues where some information that may be considered private may end up being used [69].

- **Co-operation**: Some stakeholders in companies may not be willing to provide all relevant information. Some may intentionally hide information to keep fraudulent activities hidden [69].

© University of Pretoria

Chapter 3

Literature Review: Financial industry regulation

To cope with the ever-changing financial industry, mechanisms have been put in place to regulate the financial markets. Regulation is mainly aimed at making sure that financial institutions are financially sound and can demonstrate solvency [70] and [71]. Also, regulation is aimed at preventing financial fraud and protecting consumers from unfair market practices or to compensate them in the event of default on financial institutions.

The industry regulation is almost uniform for different sectors or jurisdictions. The financial industry is broadly divided into two main sectors: banking and non-banking sectors. Regulation of the banking sector which is covered by the Basel Accords and non-banking sector is covered by the EU Life and Non-life Directives commonly know as Solvency I and II [70] and [71].

© University of Pretoria

3.1 Regulation: Banking sector

In 1988, the Basel Committee on Banking Supervision (BCBS) agreed on a set of regulatory guidelines that member states could adopt when setting up their local supervisory frameworks. This set of guidelines is called the 1988 Basel Accord or Basel I. In 2004, BCBS amended Basel I into what is now referred to as Basel II. This regulatory framework was significantly different from its predecessor as it moved from a rules-based to a principle-based regime. Even after this improvement, Basel II proved to be insufficient to prevent the 2008 Financial crisis. In 2009 further amendments were proposed to form the Basel III accord and this is being currently being implemented.

3.1.1 Basel I

The main aim for Basel I was to create a fair competitive environment for international banks and limit banks' exposure to credit risk. Though this framework took into account some risk exposure, it manly covered credit risk and not operational risk which was still considered as part of 'other' risks. The collapse of large financial institutions such as Barings Bank in the 90s is what resulted in the formulation of Basel II where operational risk got more attention.

3.1.2 Basel II

Under Basel II, operational risk was formally defined as the risk of loss due to inadequate or failed internal processes, people and systems or external events [12], [11]. Methods of calculating capital requirements specifically for operational risk were set out. The amendments were focused on mitigating likelihood of insolvency by moving from a rule based to a principle-risk approach. The framework is set out in three pillars.

© University of Pretoria

Pillar I: Minimum capital requirements

Under this pillar, banks have to calculate minimum solvency requirements mainly for credit, market and operational risk. This is an improvement from Basel I which set out capital requirements for only credit risk. There are three options for calculating the capital requirement and they all vary in complexity. These are basic indicator approach, standardised approach and advanced measurement approach. These first two use indicators to measure risk exposure and the third allows banks to create their own models to measure risk exposure and in turn calculate capital requirement. These are further discussed in section 4.2 to section 4.3.

Pillar II: Supervisory review

Not all risks are covered in pillar I and so under this pillar, banks need to assess the bank's risk profile to see if there is need for additional capital above the capital requirements set under pillar I. This gives supervisory bodies better capacity to intervene when necessary to prevent banks becoming insolvent. All risks should be identified and mitigating strategies put in place [72]. Requirements set under Pillar II are not prescriptive allowing banks to detail the risks they face individually rather than risks faced by the industry as a whole. This flexibility results in high subjectivity which may make it difficult for supervisory authorities and also result in regulatory arbitrage [73].

Pillar III: Market discipline

Market discipline has to deal with disclosure of information to the supervisory body, the market and customers. Twice a year, banks need to disclose rating gradings, risk management policies, management policies, loss experience among other information. This information is highly unique and there is need to decide what to include in these reports and what is considered confidential [73] and [72].

© University of Pretoria

3.1.3 Basel III

The BCBS in March 2016 [74] proposed an approach to calculate the capital requirement for operational risk. This is called the Standardised Measurement Approach (SMA) which is planned to be used in place of the 3 approaches set out in Basel II. This approach aims at increasing the level of homogeneity in results of capital modelling and also introduce a simpler and comparable method of setting out capital requirements [75]. Though the approach is standardised, it is still risk-sensitive by combining financial statement proxy for risk exposure and company loss data.

Business indicator component

The first component is the Business Indicator (BI). This will be used instead of the Gross Income (GI) used in BI and SA approaches. To this interest, lease and dividend component (ILDC), services component (SC) and financial component (FC) are added. ILDC is self-explanatory on what it includes. SC will include operating income, operating expenses and fees while FC includes the balance on the trading and trading books. The method utilises absolute values of each item that contributes to BI unlike with GI where negative values can reduce GI [74].

To get the Business Indicator Component, companies are divided into 'buckets' based on the size of the BI they have. The categorization is shown in Table 3.1 below

Internal loss multiplier

The internal loss multiplier (ILM) is based on the loss component (LC) which is derived from the internal loss data of the company. The bank is advised to use loss data from a 10-year period and the dataset must be of good quality. The data set must be of

© University of Pretoria

Bucket	BI range (in €billion)	BIC(in €billion)
1	0-1	0.11*BI
2	1-3	0.11+0.15*(BI-1)
3	3-10	0.41+0.19*(BI-3)
4	10-30	1.74+0.23*(BI-10)
5	>30	6.34+0.29*(BI-30)

Table 3.1: BI buckets and corresponding BIC.

a minimum of 5 years and if the company does not have the data then capital will be based on BIC only [74].

The LC is calculated using mean annual losses where larger losses have a higher weighting. This means that if a company has more large loses, the LC will be greater as well which will result in greater capital requirement [75]. The LC is calculated as follows:

$$LC = 7 * ATAL + 7 * ATAL_{for\ Losses > €10 million} + 5 * ATAL_{for\ Losses > €100 million}, \quad (3.1)$$

where $ATAL$ is the Average Annual Total Loss [74]. The ILM is defined as:

$$ILM = \ln\left(e^1 - 1 + \frac{LC}{BIC}\right). \quad (3.2)$$

The ILM is bound below by $\ln\left(e^1 - 1\right) = 0.541$ and so a company which did not have an operational risk loss needs to set aside about half of its BIC as capital. Since IML is a logarithmic function, it is increasing at a decreasing rate. Alternative adjustments for extreme events are still under consideration [76]. In Table 3.2 a summary of the ILM for different Loss Components and Business Indicator buckets is provided.

© University of Pretoria

		LC					
		0.11	0.41	1.74	6.34	26.64	...
BIC	0.11	1.00	1.69	2.86	4.08	5.50	>5.50
	0.41	0.69	1.00	1.79	2.84	4.20	>4.20
	1.74	0.58	0.67	1.00	1.68	2.83	>2.83
	6.34	0.55	0.58	0.69	1.00	1.78	>1.78
	26.64	0.54	0.55	0.58	0.67	1.00	>1.00
	...	0.54	0.54	0.54	0.54	0.54	1.00

Table 3.2: Internal Loss Multiplier for different Business Indicator buckets and Loss Components [75].

Capital requirement calculation

The capital requirement is calculated by multiplying the BIC with the ILM. For bucket 1, the capital requirement is just the BIC only because the ILM is considered to be 1. For buckets 2 to 5, the first €110 million is not multiplied by the ILM. The formula, in €, is defined as [74]

$$SMA\ Capital = \begin{cases} BIC & for\ bucket\ 1 \\ 110\ million + (BIC - 110\ million)\,.ILM & for\ buckets\ 2\ to\ 5 \end{cases} \tag{3.3}$$

This new approach removes the need for AMA approach when calculating solvency capital requirements [77] and [78]. But since it still requires loss data, companies may need to assess the loss severity and frequency distributions for internal management purposes and so LDA is still an important risk management tool.

© University of Pretoria

3.2 Regulation: Insurance sector

3.2.1 Solvency I

The Council of the European Union introduced formal regulation in the insurance industry in 1973 when the first Non-Life Directive was set out for the European Union insurance industry. This was followed up in 1979 by the first Life Directive. These were revised in the mid 1990s to cope with the changes occurring in industry. Under these directives, solvency margins were calculated using rudimentary factors and formulae. Capital was calculated purely on a mathematical basis without taking into account the business risks involved.

Minimum Solvency Margin (MSM) was calculated as:

$$MSM = 4\%Reserves + 0.3\%Cap \tag{3.4}$$

where *Reserves* are the mathematical reserves gross of reinsurance *Cap* is the capital sum at risk [73]. This formula was simple and easy to administer but did not take into account the full complexity of the risks faced by insurers such as fraud, poor corporate governance and other operational risks. This also did not provide the much-needed protection for consumers. There was lack of consistency as assets could be valued on a historical or amortisation basis, not just at the market value.

In 2002, Council of the European Union introduced more directives. This resulted in more parameters being introduced for solvency evaluation to account for the non-uniformity of the business risk. For the first time, regulation influenced the type of assets that are included in the solvency margins. There was also risk classification where there were 3 main categories — Technical, Investment and Non-technical [73].

© University of Pretoria

These directives are collectively called Solvency I.

Valuations moved from the point-in-time approach that was used under the Directives. Under the Directives, solvency needed to be demonstrated at a single point in time, i.e. on date or most recent balance sheet. Also, solvency needed to be demonstrated at every point in time hence becoming more prudent. There was a requirement to set out extra solvency margins for some business contracts such as unit-linked contracts [79]. Although there were additional parameters added to valuations, it remained very simple and easy to administer. However, it did not take into account the business risk faced by insures.

3.2.2 Solvency II

Solvency I was a rule-based regulatory framework. In 2009, the EU started drafting a better framework called Solvency II which came into full effect in 2016. The regulatory regime moved from being rule based under Solvency I to principle-risk based under Solvency II. Solvency II is significantly different from Solvency I. It is expected to provide more protection to consumers and maintain financial stability in the insurance industry [79] and [31].

Solvency II aims at assessing the risk-profiles of individual insurers. This allows insurers to assess, measure and manage their risks to comply with regulation. Valuation methods are now based only on a fair value basis for both assets and liabilities [72]. This results in higher capital requirements that will give the regulator time to intervene if there are solvency issues with a specific insurer. Higher solvency margins will allow insurers to be able to withstand unexpected losses that may arise. Solvency II has 3 main pillars that are almost similar to the ones under Basel II [70].

© University of Pretoria

Pillar I: Minimum capital requirements

Under this pillar, insurers are required to maintain solvency by making sure that they maintain the Minimum Capital Requirement (MCR) and the Solvency Capital Requirement (SCR).

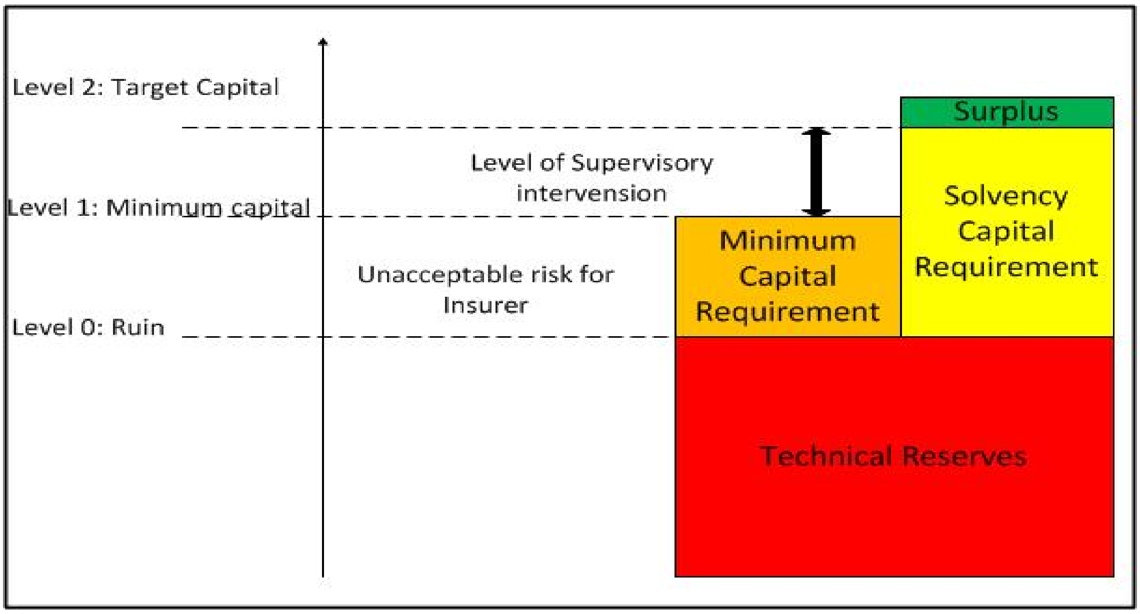


Figure 3.1: Split of a company's capital.

The SCR is the amount of capital required to maintain solvency over 1 year with a 99.5% level of confidence. The MCR is the amount of capital required to maintain the supervisory body's confidence that the insurer will be solvent over 1 year with an 85% level of confidence [80]. If the insurer does not have enough capital to reach SCR, the supervisory body will intervene as the company is considered to be in distress. If the company does not have enough capital to cover MCR, it shows the company is in financial distress and so the supervisory board has to take tougher measures to make sure the insurer does not collapse. Technical reserves is the capital that the insurer needs to meet its financial obligations. Once the reserve falls into the technical reserves, it means the insurer is insolvent and so had to stop operations [81]. Solvency II introduced capital requirements for operational risk.

© University of Pretoria

Pillar II: Supervisory review

This Pillar tends to directly and indirectly influence insurers' exposure to operational risk. This pillar focuses on the corporate governance and risk management practices an insurer has to follow. The pillar advocates for insurers to clearly set out business practices and well documented systems. Also, companies need to clearly set the scope of all risks they face, including operational risk so that they are thoroughly assessed and appropriate mitigating strategies put in place. Companies have to come up with comprehensive risk profiles that are frequently reviewed to capture any emerging risks and effectiveness of mitigating strategies in place.

Companies have to perform a risk-based internal assessment on its business. Internal models need to be set up to monitor internal processes such as operational risk management. They need to be validated and tested to check stability and sensitivity. These models will form part of the Own Risk and Solvency Assessment (ORSA). Companies need robust and accurate model to model operational risk and other risks [81].

Pillar III: Market discipline

This pillar focuses on the level of disclosure that insurers should adopt. It sets out what is required by the supervisory body and how frequent insurers have to submit documents. Insures are encouraged to state out their public disclosure and supervisory reporting policies clearly.

Insurers have to publish their solvency and financial positions and the companies' risk positions. If they do not meet the solvency requirements, they are required to disclose this information to all stakeholders. Insures have to submit detailed supervisory reports to the supervisory bodies. These reports will include information that can be

© University of Pretoria

considered to be too detailed or confidential for the public. In aim to harmonise the insurance industry, insurers are provided reporting template forms they have to fill out though they have to perform ORSA [81].

© University of Pretoria

Chapter 4

Literature review: Operational risk capital calculation approaches

This chapter covers the traditional methods of calculating capital for operational risk as specified by the Basel accords and Solvency directives.

4.1 Basic indicator approach

This is the simplest and most crude method of the three methods of calculating capital requirements. Under this method, capital reserves are calculated by taking a fixed percentage of the annual average gross operating income. In this case, gross operating income is considered to be a good representation of operational risk exposure an institution faces. The formula is:

$$Cap_{BI} = \alpha \times \frac{\sum_{i=1}^{n} max(0; I_i)}{n^*} \tag{4.1}$$

where n^* represents the total number of years in which gross income was positive[82] and [11], Cap is the capital charge and I_i is the gross income in year i. Under Basel II, n is set at 3 years and α is set at 15%. This method is considered crude because

© University of Pretoria

income received does not fully represent the operational risk exposure. Also, income may fluctuate greatly, not because of change in operational risk exposure but due to other factors.

4.2 Standardised approach

This is a special case of the basic indicator approach. Instead of taking a percentage of total gross income, capital reserves are calculated by aggregating a percentage of the gross income for each business unit. The percentage differs for different business lines. A different approach of using outstanding loans instead of gross income can be used. The formula to calculate capital is:

$$Cap_{SA} = \frac{\sum_{j=1}^{8} \beta_j \times \sum_{i=1}^{n} max(0; I_{ij})}{n^*} \tag{4.2}$$

where Cap is the capital charge and I_{ij} is the gross income in year i for business line j The rationale behind this approach is that different business line face different level of operational risk exposure. Hence, β_j will represent the risk exposure for business line j [82] and [11].

4.3 Advanced measurement approach

Internal measurement approach

This is a bottom-up approach of calculating capital charge. Operational risk exposure is divided into subcategories according to business lines and event type. This will result in cells that have a specific event occurring in a specific business line. Based on the losses that have occurred, capital charge is calculated as,

$$Cap_{ij} = EI_{ij} \times PE_{ij} \times L_{ij}, \tag{4.3}$$

© University of Pretoria

where Cap_{ij} is the capital charge for a specific business unit j towards a specific risk event i and EI_{ij} is Exposure indicator for a specific business unit j towards a specific risk event i. Different business lines have different exposures to different risk events. PE_{ij} is the probability of event i occurring in business unit j and L_{ij} is the loss amount for risk event i in business unit j [83] and [11]. The total capital charge will be the sum of the capital charge for each individual cell [11] and [82].

4.4 Operational risk measures

4.4.1 Value at risk

In order for banks and insurance companies to manage operational risk, they need to measure accurately their financial exposure. A method that is based on the frequency and loss distributions in now widely accepted as a good measure of operational risk. Once loss severity and frequency distributions are modelled using either compound distribution or EVT, companies need to find out the largest loss they can expect in a specified time period given a specific level of confidence. This is called value at risk and it is affected by three main factors: horizon, level of confidence and currency [84], [85] and [86].

Horizon - this is the time period for which the value at risk is calculated. This was set to 1 year under Basel II for banks and still is under Solvency II for insurance companies [87]. However, this has been reduced to 1 day under the new Basel III framework.

Level of confidence - internal management and regulators alike prefer this to be as close to 100% as possible. For banks this is set to 99% and 95% for insurance companies.

Currency - this is more relevant to companies that operate in multiple jurisdictions.

© University of Pretoria

All losses should be converted to a single currency, usually the one in which the headquarters offices are based, at prevailing exchange rates [11].

Mathematically, let X be the loss amounts at $(1-\alpha)\%$ confidence level. The Value at risk is defined as:

$$VaR_{\alpha}(X) = inf\left\{x \,\epsilon\, \mathbb{R}:\; P(X \leq x) \geq \alpha\right\}. \tag{4.4}$$

To calculate this value, there are three methods. The variance-covariance and historic simulation methods [88] are not used in operational risk very often because they assume that losses follow a normal distribution which is not a heavy-tailed distribution. The Monte Carlo simulation method is the common method used and data is generated using Monte Carlo simulation. The Value at risk will be the corresponding $(1-\alpha)th$ percentile [87].

The main drawback to this method was that it required a lot of computational power to perform the Monte Carlo simulations. With advances in mathematical and statistical programming software, the amount of time spent simulating data has significantly been reduced.

4.4.2 Expected shortfall

This is also called Conditional VaR, Tail VaR or Average VaR. This is considered a better alternative to VaR because it is coherent. A coherent risk measure meets the following characteristics: translation, in-variance, positive homogeneity, sub-additivity and monotonicity [89].

Expected shortfall is the expected loss that may occur given that the loss is greater than the value at risk. Instead of giving the maximum possible loss, expected shortfall gives an expected loss amount at a given level of confidence and it is defined as:

© University of Pretoria

$$ES_\alpha(X) = E\left[X \mid X > VaR_\alpha(X)\right] \tag{4.5}$$

4.5 Level of completeness in operational risk modelling

For risk management purposes, understanding and being able to model the distribution of larger losses is very important. Because of this, truncated distributions and the POT method are the more commonly used method to model losses [45]. The accuracy of this method depends on the level of the threshold set. EVT assumes that the distribution of extreme values is asymptotic i.e. a limiting distribution and so if the threshold is set too low, this assumption will no longer hold. If the threshold is instead set too high, there will be fewer data which will result in higher volatility in parameter estimates [90].

Several methods have been developed on how to select the threshold value. The most common one is the visual method where the point on the *qq* plot where the plot becomes linear is considered to be the threshold point [91]. Another method is to have a range of possible values for the threshold. Then fit a Gaussian distribution to a sample and the correct threshold will be the one that satisfies the test for Normality [92].

This selection of the threshold level is highly subjective. In the banking sector, Basel II proposes a data collection threshold of £10,000. This value is an estimate set by the BCSC and it does not take into account the actual losses a bank incurs. This problem of setting a minimum threshold value is not unique to operational risk modelling. It is also encountered in seismology.

© University of Pretoria

4.5.1 Earthquake frequency-magnitude distribution

Seismology is the precursor to earthquake detectability. The distribution of earthquake frequencies and their magnitudes can be used to predict future earthquakes and have mitigating measures against the impact of possible future earthquakes. The Gutenberg-Richter law [93] is the common technique used to describe the relationship between earthquake magnitude and frequency due to it being ubiquitous. The law states that the logarithm of the number of earthquakes of a specified magnitude and the magnitude of the earthquake has a linear relationship [94], [13]. Let M be the magnitude of an earthquake and N is the number of earthquakes of magnitude M and above. Let a and b be parameters describing the relationship. Then,

$$log\, N = a - b(M - M_c). \tag{4.6}$$

This is a linear relationship with b as the slope. The slope, b represents the ratio of the number of large earthquakes to smaller ones in terms of magnitude. The value for b will vary for different locations because different locations have different seismicity. This relationship between the frequency and magnitude of earthquakes can be modified to form an exponential distribution defined as follows,

$$f_M^*(m) = \beta e^{-\beta(m - M_c)} \qquad m \geq M_c, \tag{4.7}$$

where $\beta = b \ln 10$. M_c is called the level of completeness. There is a negative correlation between the logarithm of number of earthquakes and magnitude. The 'high frequency - low impact ' and 'low frequency - high impact' property that is found in operational risk is also shown in seismology.

© University of Pretoria

4.5.2 Level of completeness in seismology

Not all natural earthquakes that occur are detected and not all earthquakes detected are natural earthquakes. At lower magnitude levels, some earthquakes will not be detected due to several possible reasons. Seismographs are not 100% perfect instruments and so data that is collected is not complete. However, as the magnitude increases, the earthquakes get harder to miss and so there is a magnitude level where all earthquakes of this magnitude and above are detected. This means the dataset is complete for earthquakes above this point. This point is called the level of completeness m_c.

As shown in Fig. 4.1 for earthquakes above the level of completeness, all the earthquakes are observed and recorded and the distribution follows the Gutenberg-Richter formula. However, for earthquakes below the level of completeness, the actual distribution departs from the Gutenberg-Richter law. This is because not all earthquakes are observed at lower magnitudes.Equation (2) can be transformed into the probability distribution of the earthquake magnitude with the form of an exponential distribution.

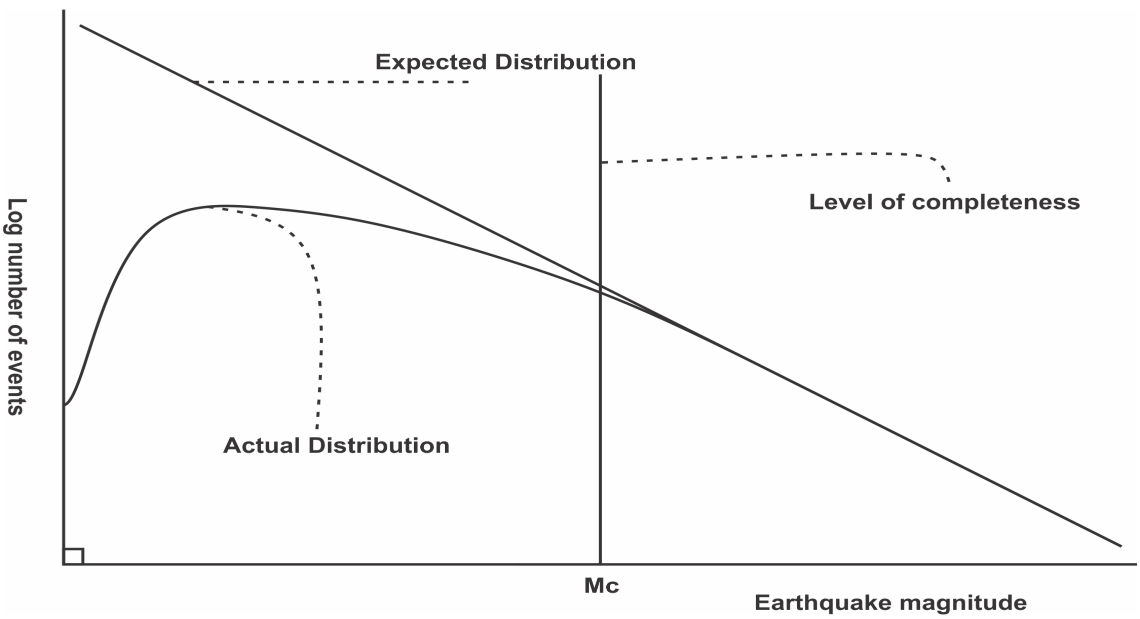


Figure 4.1: Expected vs Observed distribution showing how the actual distribution depart from the expected distribution at lower end. [17]

© University of Pretoria

The level of completeness is not known in advance and so techniques have been developed to estimate it. The most common method is the Maximum Curvature technique (MAXC) [95]. The level of completeness is set as the magnitude that gives the highest value of the frequency-magnitude distribution. The second method is the Median Based analysis of slope segment (MBASS). This is a non-parametric method based of the Wilcoxon sum rank test. The method aims to find the points of discontinuity of the frequency-magnitude distribution and the point where there is the main discontinuity is considered the level of completeness [96]. Other estimation methods developed include the Goodness-of-fit test [95], M_c by b-value stability [97], M_c from the Entire Magnitude Range [98] and [99].

4.6 Summary

The financial environment has become more sophisticated. As much as the monitoring systems have improved in line, there are still concerns on the financial soundness of financial institutions. More financial products are being produced and institutions are more connected and so the systemic risk is high. There is increased competition and so institutions need to thoroughly analyze their risks and set appropriate risk mitigation measures and enough capital. Data mining and machine learning fraud detection techniques can be used to improve risk management by making loss datasets more comprehensive by identifying even the hidden losses.

The choice of threshold has a significant impact on the parameter estimates of distribution and so by introducing a detection function, this can be remedied. Regulatory regimes advise on levels of thresholds but it lies with the companies to set more appropriate thresholds specific to that company. The nature of operational risks a company faces can make it complicated in setting out the correct threshold values to make the

© University of Pretoria

datasets complete. Even this was initially set out by regulatory bodies which are now moving towards formulae-based approaches, it is still important to perform losses analysis for internal risk management purposes for financial institutions.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Chapter 5

Methodology

5.1 Introduction

Small losses are considered to be of little consequence to the profitability and solvency of a company as these are regarded predictable and expected. And so operational risk modelling techniques utilize truncated distributions to model the upper end of the loss distribution where there are large unexpected losses. In order to assess the distribution of smaller losses, a new distribution that allows for random deletions is constructed and fitted to the data set instead of a truncated distribution.

In this chapter, the proposed 3-parameter Gamma distribution will be defined and the parameter estimation with be explained.

5.2 Operational risk loss model

Operational risk losses are usually modelled by fat-tailed distributions. The methodology explained above assumes an exponential distribution. Pareto distribution is fat-tailed and is used to model operational risks. For the method to be applied, the Pareto

© University of Pretoria

distribution needs to be converted to an exponential distribution as follows. Let X be the Pareto distributed with parameter β. A new variable is defined as $h(x) = Y = ln\left(\frac{X}{x_{min}}\right)$. The distribution for Y is:

$$\begin{aligned} f_Y(y) &= f_X\left(h^{-1}(y)\right)\frac{d}{dy}h^{-1}(y) \\ &= \frac{\beta(x_{min})^{\beta}}{(x_{min}e^{y})^{\beta+1}}(x_{min}e^{y}) \\ &= \beta e^{-y\beta}. \end{aligned} \tag{5.1}$$

The method will use the transformation described above. Aggregate losses however will be based on the results that will be transformed back using Eq. 5.1.

Initially, there is the Gutenberg-Richter equation [94]

$$log\, N = a - bX. \tag{5.2}$$

X is the singly reported loss amount and N is the number of losses of loss amount X. a and b are the parameters that explain the relationship between the logarithm of the number of losses and loss amount. From Eq. 5.2 it follows that the distribution of losses can be transformed to an exponential distribution as shown below.

$$\begin{aligned} logN &= a - bM \\ N &= 10^{a-bM} \\ &= 10^{a} * 10^{-bM} \\ &= 10^{a} * (10^{b})^{-M} \\ &= 10^{a} * (e^{\beta})^{-M} \\ &\propto e^{-\beta M} \end{aligned}$$

After normalization, the distribution will be defined as

$$f_X(x) = \beta e^{-\beta(x-x_0)} \qquad for\ 0 \leq x < \infty, \tag{5.3}$$

© University of Pretoria

where x_0 is the minimum loss recorded and $\beta = b\ln(10)$ is the distribution parameter. This distribution assumes that the dataset available is complete. In practice, some small losses that occur to random errors or ones that are purposefully hidden in fraudulent activities do not get recorded and so the dataset is incomplete. However, all losses that are greater than the level of completeness will be detected. The challenge is to identify this level of completeness since it is not known in advance.

5.2.1 Random deletion

Loss data is usually missing for small values. Some losses are so small that they are considered immaterial and so are not recorded. Some are randomly missed due to human error. Some losses are intentionally hidden when there are fraudulent activities occurring. These losses are not necessarily small as seen in recent years, when big banks like Barings collapsed due to rogue trading that went hidden and went unnoticed.

Because the level of completeness is unknown, a probability distribution of data points that will be included in the data set is defined. Mainly, a Gaussian and an angular distribution are applied [100]. To allow the distribution to assume different lower tail shape that is dependent on the data set, a polynomial function is selected [17]. Let $p_c(x)$ be the probability that a data point is included in the dataset and is defined as

$$p_c(x) \propto (x - \phi)^{\alpha - 1} \qquad \phi < x \leq x_{max},\ \alpha > 0, \tag{5.4}$$

where ϕ is the location parameter and α is the shape parameter.

© University of Pretoria

5.2.2 Apparent distribution

To get the apparent distribution of the losses, multiply the loss distribution with the probability of inclusion

$$f_Y(x) = Kp_c(x)f_X(x), \tag{5.5}$$

where K is the distribution normalizing constant.

$$\begin{aligned}
f_Y(x) &= Kf_X(x)\, p_C(x) && (5.6)\\
&= K\beta e^{-\beta(x)} (x-\phi)^{\alpha-1} && (5.7)\\
&= K\beta e^{-\beta(x)} (x-\phi)^{\alpha-1} e^{-\beta\phi} e^{\beta\phi} && (5.8)\\
&= K\beta e^{-\beta(x-\phi)} (x-\phi)^{\alpha-1} e^{-\beta\phi} && (5.9)\\
&= \frac{K\beta e^{-\beta(x-\phi)} (x-\phi)^{\alpha-1} e^{-\beta\phi}\Gamma(\alpha)\beta^{\alpha-1}}{\Gamma(\alpha)\beta^{\alpha-1}} && (5.10)\\
&= \frac{\beta^{\alpha} (x-\phi)^{\alpha-1} e^{-\beta(x-\phi)}}{\Gamma(\alpha)} \qquad (\alpha > 0,\ \beta > 0,\ x > \phi)\,. && (5.11)
\end{aligned}$$

5.3 Parameter estimation

To estimate the parameters $\boldsymbol{\theta} = (\alpha,\ \beta,\ \phi)$ of the apparent distribution, the method of moments and the maximum likelihood estimation methods are used.

5.3.1 Method of Moments

The sample central moments about the mean are defined as [101]

$$\hat{\mu}_k = \frac{\sum_n^{i=1}(X_i - \bar{X})^k}{n}. \tag{5.12}$$

The population central moments about the mean are defined as [101]

© University of Pretoria

$$\mu_k = E((Y-\mu_Y)^r) = \int_{-\infty}^{\infty} (y-\mu_Y)^r f_y(y)dy. \tag{5.13}$$

Since there are three unknown parameters, a set of three equations is set by equating the first 3 sample moments to the first 3 population moments.

$$\begin{cases} \hat{\mu_1} &= \mu_1(\boldsymbol{\theta}) = \frac{\hat{\alpha}}{\hat{\beta}} + \hat{\phi} \\ \hat{\mu_2} &= \mu_2(\boldsymbol{\theta}) = \frac{\hat{\alpha}}{\hat{\beta}^2} \\ \hat{\mu_3} &= \mu_3(\boldsymbol{\theta}) = \frac{2\hat{\alpha}}{\hat{\beta}^3} \end{cases} \tag{5.14}$$

and the solution to this equation is

$$\begin{cases} \hat{\alpha} &= \frac{4\hat{\mu_2}^3}{\hat{\mu_3}^2} \\ \hat{\beta} &= \frac{2\hat{\mu_2}^2}{\hat{\mu_3}} \\ \hat{\phi} &= \hat{\mu_1} - \frac{2\hat{\mu_2}^2}{\hat{\mu_3}}. \end{cases} \tag{5.15}$$

A bootstrapping strategy is used to calculate the variance of the estimates.

5.3.2 Maximum Likelihood Estimation

Under this method, the log likelihood function is developed by taking the logarithm of the product of the probability density function of n identical and independent random variables and shown in Eq. 5.16.

$$\begin{aligned} L(\boldsymbol{\theta}|\boldsymbol{y}) &= \prod_{i=0}^{n} \frac{\beta^\alpha (y_i-\phi)^{\alpha-1} e^{-\beta(y_i-\phi)}}{\Gamma(\alpha)} \\ &= \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^n \left(\prod_{i=1}^{n}(y_i-\phi)^{\alpha-1}\right) e^{-\beta\sum_{i=0}^{n}(y_i-\phi)} \\ \ln L(\boldsymbol{\theta}|\boldsymbol{y}) &= n\ln\left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right) + (\alpha-1)\sum_{i=1}^{n}\ln(y_i-\phi) - \beta\sum_{i=1}^{n}(y_i-\phi) \end{aligned} \tag{5.16}$$

Taking partial derivatives with respect to each of the three parameters and equating them to zero forms a set of equations.

© University of Pretoria

$$\begin{cases} \frac{\partial}{\partial\beta}\ln L\left(\boldsymbol{\theta}|\boldsymbol{y}\right) &= \frac{n\hat{\alpha}}{\hat{\beta}} - \sum_{i=0}^{n}(y_i - \hat{\phi}) = 0 \\ \frac{\partial}{\partial\alpha}\ln L\left(\boldsymbol{\theta}|\boldsymbol{y}\right) &= \frac{n\hat{\beta}^{\hat{\alpha}}\ln\hat{\beta}}{\hat{\beta}^{\hat{\alpha}}} - \frac{n\Gamma\prime(\hat{\alpha})}{\Gamma(\hat{\alpha})} + \sum_{i=1}^{n}\ln(y_i - \hat{\phi}) = 0 \\ \frac{\partial}{\partial\phi}\ln L\left(\boldsymbol{\theta}|\boldsymbol{y}\right) &= (1-\hat{\alpha})\sum_{i=1}^{n}(y_i - \hat{\phi})^{-1} - n\hat{\beta} = 0 \end{cases}$$

This can be further simplified to

$$\begin{cases} \frac{\hat{\alpha}}{\hat{\beta}} - \overline{(y-\hat{\phi})} = 0 \\ \ln\hat{\beta} - \psi(\hat{\alpha}) - \overline{\ln(y-\hat{\phi})} = 0 \\ (1-\hat{\alpha})\overline{(y-\hat{\phi})^{-1}} - \hat{\beta} = 0, \end{cases} \tag{5.17}$$

where $\overline{(y-\hat{\phi})} = \frac{\sum_{i=0}^{n}(y_i-\hat{\phi})}{n}$, $\overline{\ln(y-\hat{\phi})} = \frac{\sum_{i=1}^{n}\ln(y_i-\hat{\phi})}{n}$, $\overline{(y-\hat{\phi})^{-1}} = \frac{\sum_{i=1}^{n}(y_i-\hat{\phi})^{-1}}{n}$ and *Psi*(Digamma) function $\psi(\alpha) = \frac{\Gamma\prime(\alpha)}{\Gamma(\alpha)}$. The solution of this system of equation is obtained through iterations since i does not have a closed solution.

The *Psi* function is approximated by [102]:

$$\psi'(\alpha) \approxeq \frac{1}{\alpha} + \frac{1}{2\alpha^2} + \frac{1}{6\alpha^3}. \tag{5.18}$$

Based on the approximation in in Eq. 5.18, the variance of the estimates can can be calculated as follows [103],

$$\begin{cases} Var(\hat{\alpha}) &\approxeq \frac{6\hat{\alpha}^3}{n} \\ Var(\hat{\beta}) &\approxeq \frac{3\hat{\alpha}}{n\hat{\beta}^2} \\ Var(\hat{\phi}) &\approxeq \frac{3\hat{\alpha}^3\hat{\beta}^2}{2n}. \end{cases} \tag{5.19}$$

© University of Pretoria

Chapter 6

Results

In this chapter, data simulation exercise carried out is discussed. A complete dataset is populated and incomplete datasets are obtained by deleting values from the original dataset. Parameter estimates are obtained and the theoretical distribution is compared to the datasets.

6.1 Simulated data set

To assess the model, a data simulation exercise to replicate a loss amount dataset was carried out. A complete dataset, to represent an ideal situation, and other datasets that had incomplete data points, to represent real life experience, were simulated. The simulated datasets were fitted to the distribution defined in Eq. 5.5 in order to assess the how β estimates will vary over different levels of deletion.

6.1.1 Complete dataset

Data is simulated from an exponential distribution with mean of 10. By definition of the mean of an exponential distribution, the parameter is 0.1. 50 000 data points are

© University of Pretoria

generated from this distribution. Data is then scaled to fit a range of 1 to 7 to constrain the range of values for analysis. The computer program dictated the range of values since it is created for earthquake magnitudes. Let X_i be the i^{th} observation of the original dataset. The following translation is done to scale the data,

$$Y_i = \frac{X_i - a}{b}. \tag{6.1}$$

The parameters a and b are obtained from solving the following system of linear equations:

$$A = \frac{X_{min} - a}{b} \tag{6.2}$$

$$B = \frac{X_{max} - a}{b} \tag{6.3}$$

where $X_{min} = 0$ and $X_{max} = 117.9$ based on the simulated values. This gives the values of a = -19.65 and b = 19.65. By transforming Eq 5.3 to the new distribution

$$f_Y(y) = e^{-a\beta}(b\beta)e^{-(b\beta)y} \qquad A \leq y \leq B, \tag{6.4}$$

where in this case $A = 1$ and $B = 7$. The dataset is fitted to the derived 3-parameter Gamma distribution

$$f_Y(x) = \frac{\beta^{*\alpha}\,(y-\phi)^{\alpha-1}\,e^{-\beta^*(y-\phi)}}{\Gamma(\alpha)} \qquad (\alpha > 0,\, \beta^* > 0,\, y > \phi)\,, \tag{6.5}$$

where $\beta^* = \frac{1}{b\beta} = \beta_{POC}$

α_{POC}, β_{POC} and ϕ_{POC} are the parameters that explain the distribution in Eq. 6.5 for data has been transformed. From the results shown in Table 6.1, the MME method produced estimates that are close to the true values, but with greater variability compared to MLE method estimates. The contrast between the fitted distribution and the simulated data can be observed in Fig. 6.1 for the log number of losses and Fig. 6.2 for

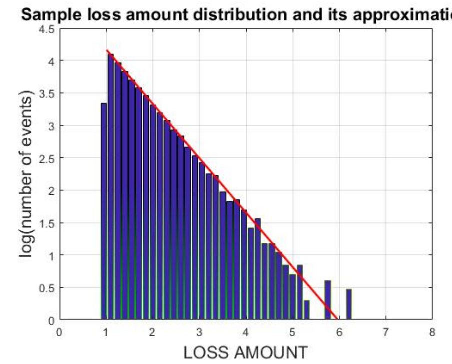
© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

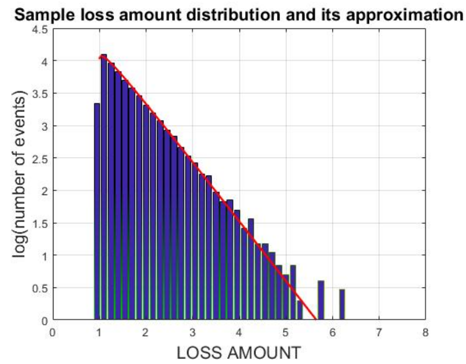
Parameter		Method of Moments		Maximum likelihood	
	Parameter value	estimate	st dv	estimate	st dv
α_{POC}	1	0.972	0.035	1.121	0.001
β_{POC}	0.5089	0.517	0.011	0.465	0.002
ϕ_{POC}	1	1.008	0.010	0.99	0.000

Table 6.1: Parameters of probability of completeness (PoC) for complete dataset.

the actual loss distribution.

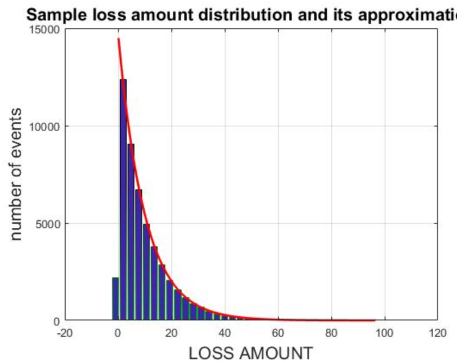


(a) method of moments

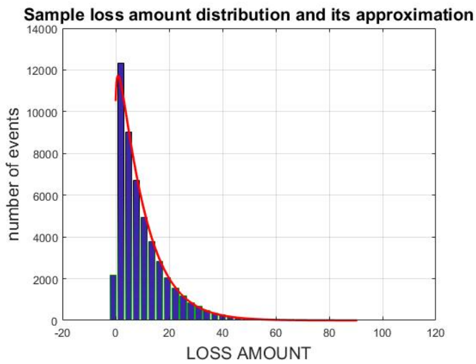


(b) maximum likelihood

Figure 6.1: Log Number of losses for complete dataset



(a) Method of moments



(b) maximum likelihood

Figure 6.2: Histograms of the complete dataset

For the complete dataset, the estimate for α is very close to 1 for the MME method. This means that the probability of inclusion is 1 for all values of y. By looking at the plots, the MME estimates give a straight line for the log-number of losses and the MLE

© University of Pretoria

estimates show better fit at the lower tail of the loss distribution.

6.1.2 Incomplete datasets

The deletion probability distribution is defined as

$$p_c(x) = \frac{\alpha(x-\phi)^{\alpha-1}}{(7-\phi)^{\alpha}-(1-\phi)^{\alpha}} \qquad 1 \leq x \leq 7. \tag{6.6}$$

which gives the following distribution function,

$$P_c(x) = \frac{(x-\phi)^{\alpha}-(1-\phi)^{\alpha}}{(7-\phi)^{\alpha}-(1-\phi)^{\alpha}} \qquad \alpha \neq 1 \tag{6.7}$$

The distribution function is used to randomly delete values from the original dataset. The deleted values will correspond to the loss amounts that are expected to be missing in real life datasets. These losses will not be observed as they will randomly be left out or intentionally hidden or too small to be considered of material value to an institution. For different values of α there will be a different dataset. A summary of the estimates for the parameter α is provided in Table 6.2:

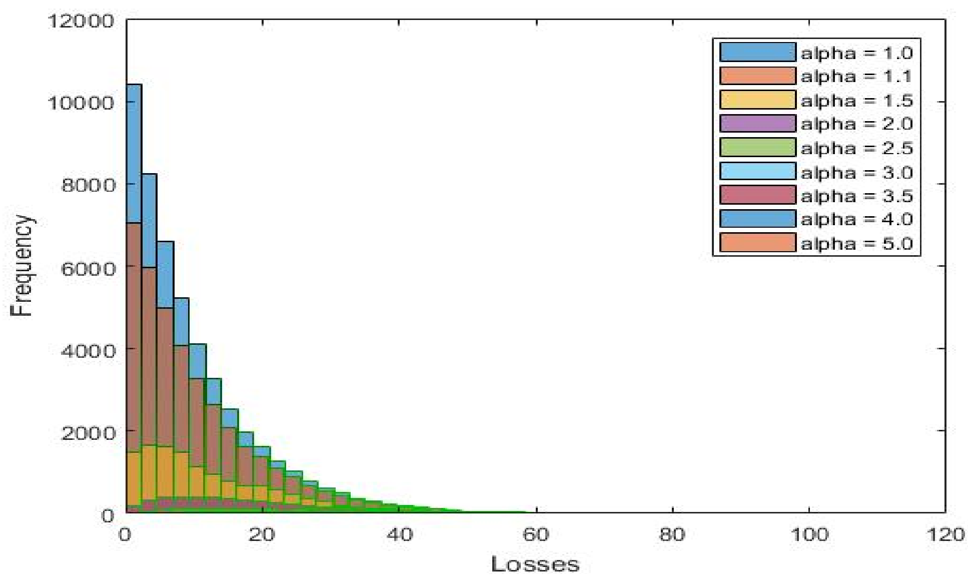


Figure 6.3: Comparative histograms for aggregate losses for different values of α

In the 6.3, the frequency of smaller losses decreases with increasing value of α as more

© University of Pretoria

losses are deleted from the dataset. The level of α shows how much losses go unnoticed and so this parameter is important to a company. Below in Table 6.3 the estimates for α and the effect they have on the dataset are set. As α increases, the dataset becomes smaller and the mean increases. The mean increases due to the reduction in the number of smaller losses. The parameter estimates depart from the true values and volatility increases as the dataset size decrease.

	Method of Moments		Maximum likelihood			
True Parameter	estimate	st dv	estimate	st dv	Dataset size	Mean
1	0.972	0.035	1.121	0.001	50 000	9.953
1.1	0.95	0.047	0.952	0.000	38 338	10,8096
1.5	1.189	0.087	0.771	0.002	13 333	13.651
2	1.66	0.165	1.499	0.014	4 784	18.883
2.5	2.010	0.295	2.237	0.076	1 871	23.8448
3	2.049	0.368	2.603	0.164	827	29.4736
3.5	2.349	0.665	3.046	0.327	404	34.4144
4	3.227	2.499	3.363	0.543	214	39.7561
5	8.339	56.328	2.806	0.569	76	53.4908

Table 6.2: Estimates for α for different data set simulated

Table 6.2 below shows the parameter estimates fitted to the 3 -parameter Gamma distribution developed in Chapter 5 where for each α value, indicates the level of deletion carried out on the initial complete data set. The MLE method provided more stable parameter estimates even at a high level of α compared to the MME estimates.

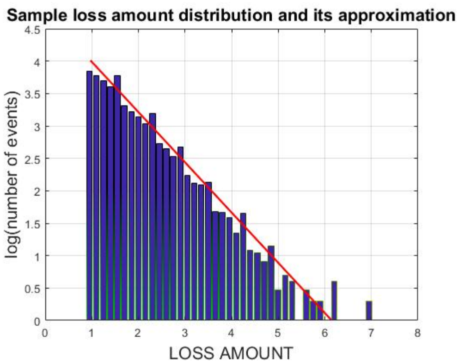
Graphs for log number of losses

In Fig. 6.4 to 6.12, histograms of the datasets (bars) and the theoretical distribution (line) are plotted.

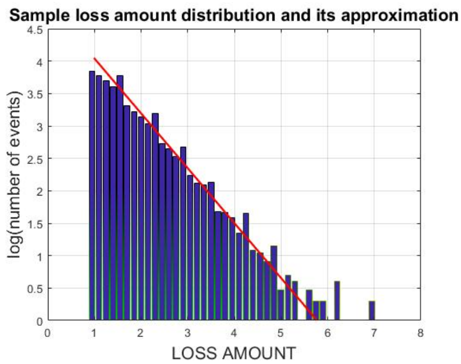
© University of Pretoria

Set with simulated data α	Method of Moments			Maximum likelihood		
	α	β	γ	α	β	γ
1	0.972	0.517	1.008	1.121	0.465	0.99
1.1	0.95	0.561	0.97	0.952	0.51	0.99
1.5	1.189	0.571	0.972	0.771	0.662	0.99
2	1.66	0.58	0.963	1.499	0.638	0.97
2.5	2.01	0.621	0.987	2.237	0.588	0.92
3	2.049	0.679	1.083	2.603	0.593	0.93
3.5	2.349	0.701	1.093	3.046	0.604	0.9
4	3.227	0.645	0.856	3.363	0.626	0.83
5	8.339	0.534	-1.036	2.86	0.959	0.73

Table 6.3: Parameter estimates for different level of random deletion

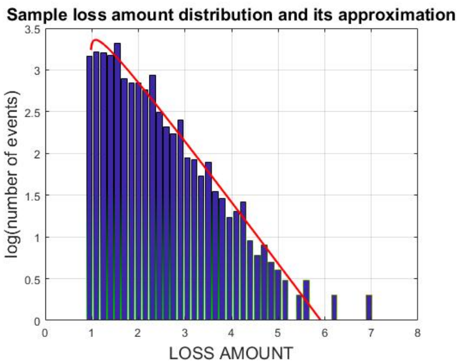


(a) Method of moments



(b) Maximum likelihood

Figure 6.4: Log Number of losses for incomplete dataset $\alpha = 1.1$

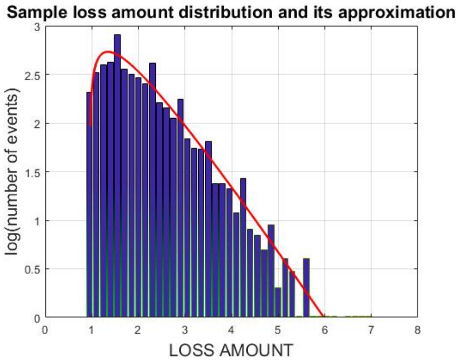


(a) Method of moments

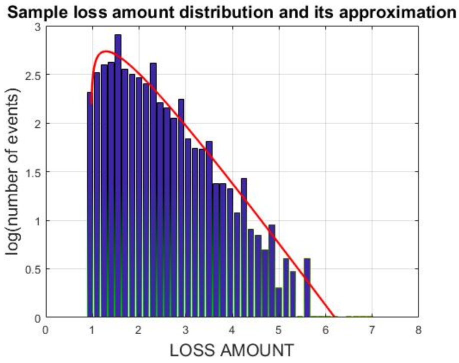
(b) Maximum likelihood

Figure 6.5: Log Number of losses for incomplete dataset 1.5

© University of Pretoria

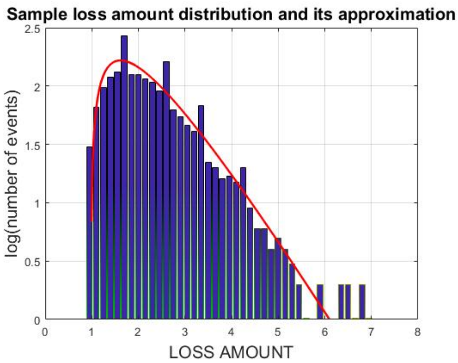


(a) *Method of moments*

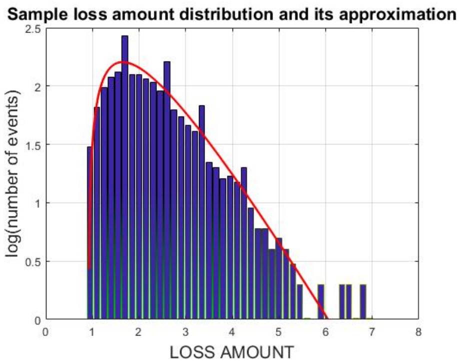


(b) *Maximum likelihood*

Figure 6.6: Log Number of losses for incomplete dataset for $\alpha = 2$

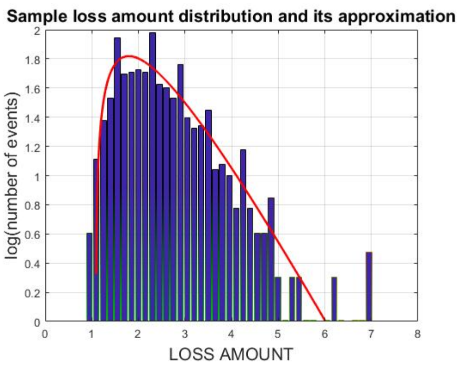


(a) *Method of moments*

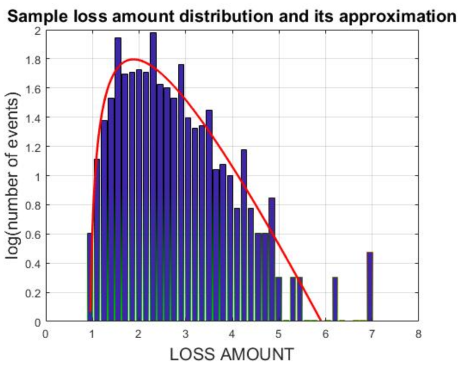


(b) *Maximum likelihood*

Figure 6.7: Log Number of losses for incomplete dataset $\alpha = 2.5$



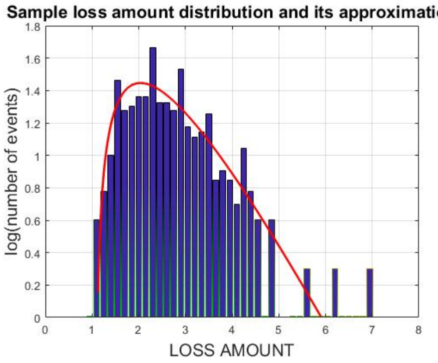
(a) *Method of moments*



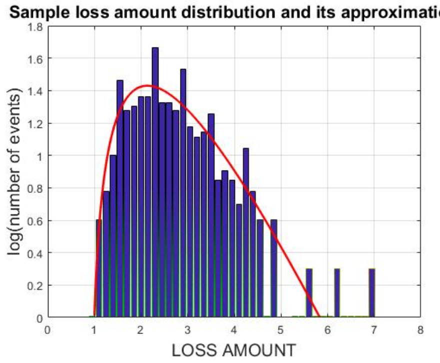
(b) *Maximum likelihood*

Figure 6.8: Log Number of losses for incomplete dataset $\alpha = 3$

© University of Pretoria

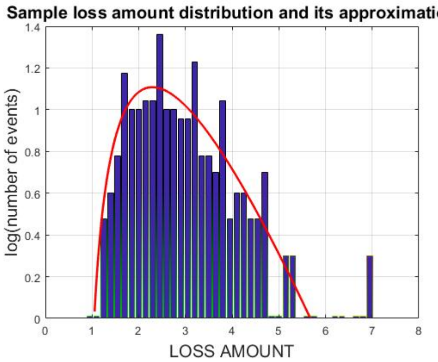


(a) *Method of moments*

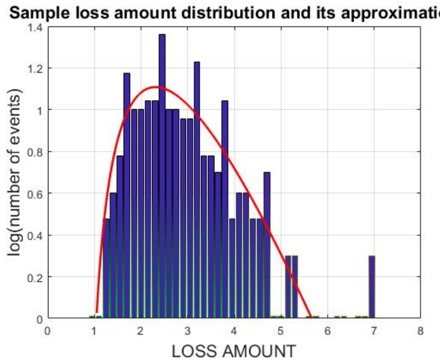


(b) *Maximum likelihood*

Figure 6.9: Log Number of losses for incomplete dataset $\alpha = 3.5$



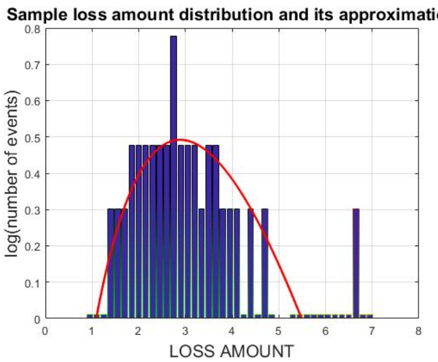
(a) *Method of moments*



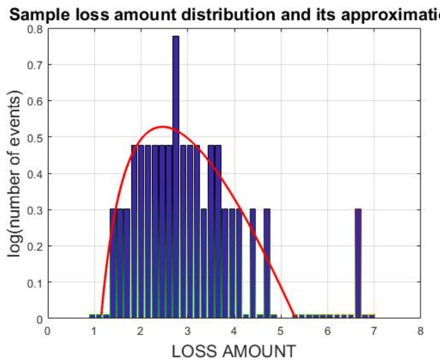
(b) *Maximum likelihood*

Figure 6.10: Log Number of losses for incomplete dataset $\alpha = 4$

Figure 6.11: Mean and Standard deviation



(a) *Method of moments*



(b) *Maximum likelihood*

Figure 6.12: Log Number of losses for incomplete dataset $\alpha = 5$

As α increases, the size of the dataset decreases and the parameter estimates for α differ significantly with higher variability. On the graphs, the bars represent the log-

© University of Pretoria

number of losses and the line represents the theoretical distribution. For $\alpha = 1$, the theoretical distribution is a straight line with a negative slope and as the theoretical Gutenberg-Richter model. As α increase, data is lost at the lower tail and the theoretical distribution curves downwards to take the deletions into account.

6.2 Application to real-life dataset

The methodology discussed in this paper is applied to a loss dataset published by Willis Towers Re. This dataset is made up of 1,078 operational risk losses publicly reported and it covers some of the notable financial fraud cases discussed in Section 2.2.1. These losses are subdivided into different loss types and the business category they occurred in. The losses are denoted in dollar-amount and inflation-adjusted public losses recorded for the period of 1975 to 2013. Some characteristics of the dataset are set out in Table 6.4 below. The estimates for α are both below 1 meaning that the detection function into a rational function which curves upwards. This gives very lower probability of detection to small losses. The probabilities increase at a significantly higher rate as loss amounts increase. The results show that the MLE estimates

Mean	US$123,716,991.31 (R1.66 billion)
Standard Deviation	US$267,644,665.25 (R3.59 billion)
Maximum	US$1,912,410,826.00 (R25,64 billion)

Table 6.4: Mean and Standard deviation of real-life dataset losses.

model the data better compared to the MME. The dataset comprised of losses of different types and so this resulted in gaps in between and a long tail. Subdividing the data result in fewer data points in each homogeneous class. The parameter estimates are in Table 6.5. The values for α and ϕ are comparable for the two methods but the same is not observed for β. The MLE method had a very high β value compared to

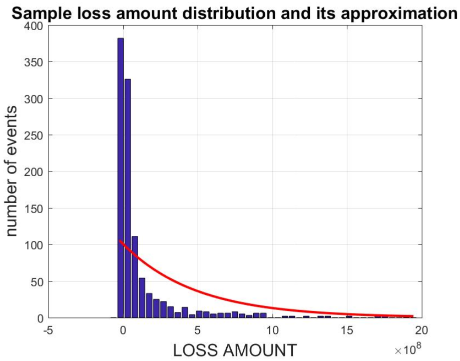
© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

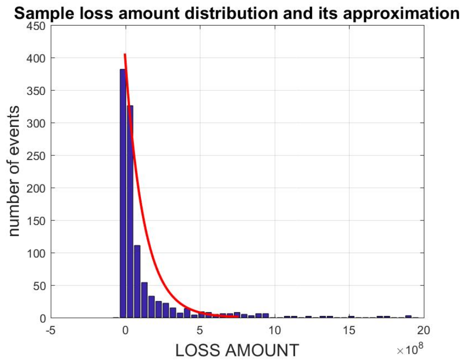
MME method. By looking at the graphs in Fig. 6.13, MLE method models the lower tail better and poorly at upper tail while the MME models the lower tail poorly but well at upper tail and this is evident in Fig. 6.14 as well. Refer to the appendix to for the results of fitted distribution on the different business lines and loss types.

Parameter	Method of Moments		Maximum likelihood	
	estimate	st dv	estimate	st dv
α_{POC}	0.3	0.031	0.33	0.008
β_{POC}	1.532	0.101	0.398	0.009
ϕ_{POC}	0.928	0.026	0.99	0.000

Table 6.5: Parameters of probability of completeness (PoC) for Willis Re dataset.

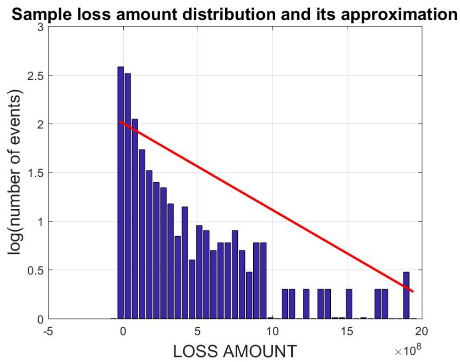


(a) Method of moments

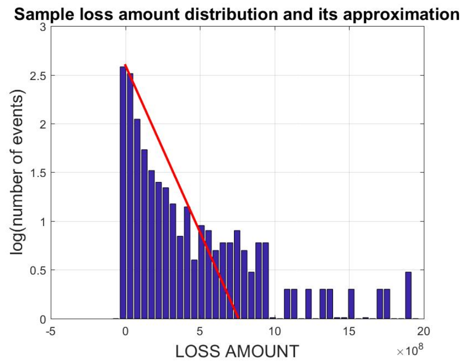


(b) Maximum likelihood

Figure 6.13: Number of losses in WillisRe dataset.



(a) Method of moments



(b) Maximum likelihood

Figure 6.14: Log number of losses in WillisRe dataset.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Chapter 7

Analysis of results

7.1 Interpretation of results

The aim of the research is to set a loss distribution that does not have explicit assumptions about the level of completeness. This loss distribution model takes into account all losses recorded irrespective of the size of the loss size. By introducing the detection probability that is a polynomial function with the polynomial degree as a parameter, the behaviour of the lower tail can be assessed.

The range of losses institutions face vary from institution to institution and so the dataset profiles will be different as well. Also based on the type on the institution, security measure and reporting standards, the level of α defined in Eq. 5.3 will vary. The value of α can give an indication on an institution's data collection and reporting practices. A high value for α can be a result of a highly incomplete dataset. This in turn gives an indication of the state of an institution's reporting standards. If the standards are not stringent enough, losses may go unnoticed or unreported because they will be hidden. Analysing the level of α over a period of time and seeing how it fluctuates can help an institution identify when reporting standards need to be adjusted.

© University of Pretoria

Eq. 5.3 provides the ideal distribution of log-losses that follow an exponential distribution. The distribution parameter, β, cannot be accurately estimated if the loss dataset is incomplete. As shown in Fig. 7.1 the actual losses would not be providing the true losses that were incurred. The losses in area A represent the ones that go undetected but may have occurred.

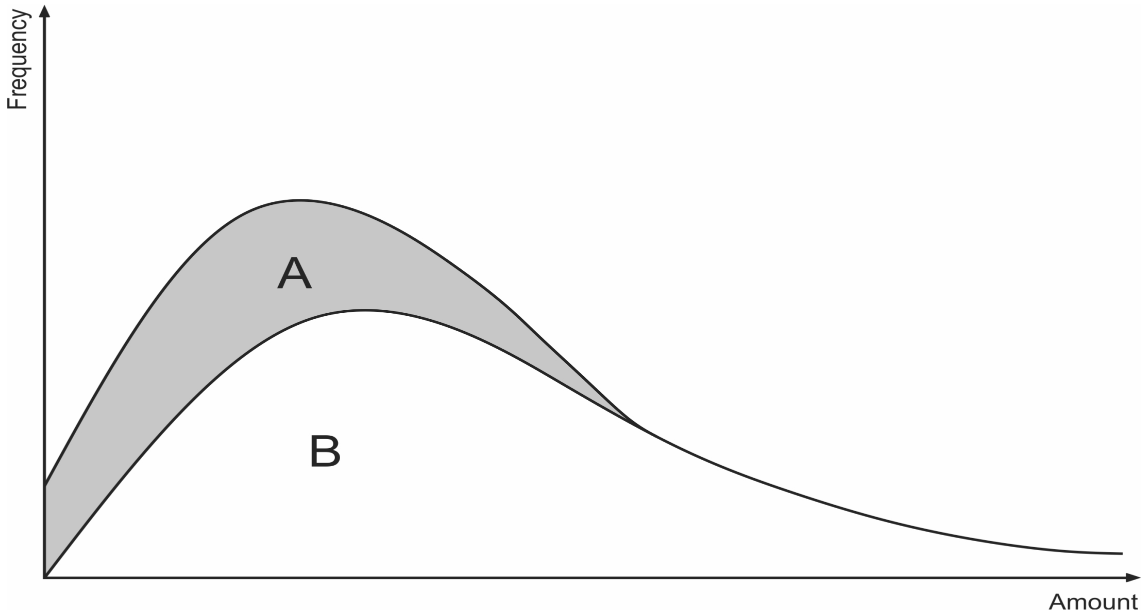


Figure 7.1: A represent the undetected losses and B represent the detected losses.

However, by introducing a detection probability, an estimate for the true distribution can be calculated. The values of β are shown in Fig. 7.2. The β estimates are stable

α	1.0	1.1	1.5	2.0	2.5	3.0	3.5	4.0	5.0
MME estimate for β	0.0994	0.0933	0.0917	0.0903	0.0878	0.0771	0.0747	0.0828	0.1117
MLE estimate for β	0.1105	0.1021	0.0791	0.0821	0.0928	0.0883	0.0867	0.0853	0.0622

Table 7.1: Estimates for the β of the underlying data for different levels of deletion.

and close to the true value of $\beta = 0.1$. This goes to show that even if there is data missing, the parameter β for the true original distribution can be found. This goes to show how even in an institution where reporting standards are less strict or if losses are are intentionally hidden, there is a way to infer on the expected experience. For an α of 4 or 5, it represents an institution that has many losses that go unreported. Even

© University of Pretoria

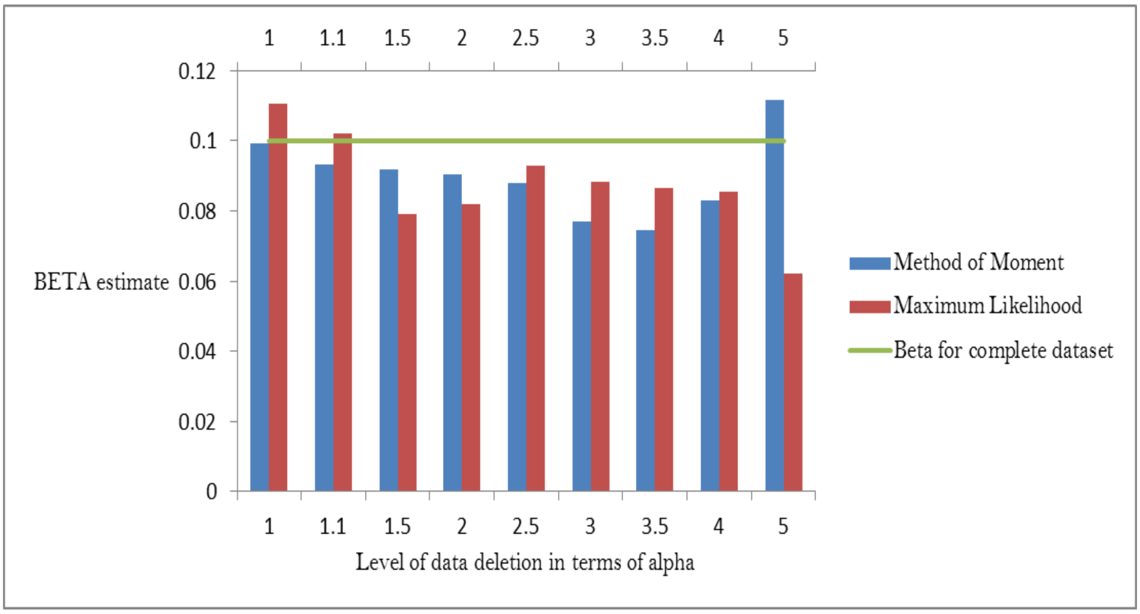


Figure 7.2: The value of beta for different estimation methods relative to the theoretical value.

at this high level of losses not getting noticed, the estimate for β is still close to the true parameter value of the underlying distribution.

Financial institutions have measures and checks put in place by local governments or by local regulators to make sure that customers are protected and employees are well trained and trustworthy enough to working in the industry. However, these measures are not fool proof and so modelling fraudulent activities can also help in setting fraud prevention measures. The proposed method can be used for this purpose. Using AMA solvency capital calculations methods, institutions can use frequency-severity analysis. An institution can incorporate a detection probability to its severity distribution and assess the impact on the unreported losses on the capital charge.

For internal management purposes, there is need to assess how internal losses are distributed and to make all losses that are incurred are identified, investigated and reported appropriately. The results will give an indication of the internal management processes when it comes to loss detection and fraud prevention. For example, if there is a significant increase in the value of α, this may indicate that more smaller losses are

© University of Pretoria

going unnoticed and so there is need to adjust the internal reporting standards or further investigate to make sure that no losses occurred or if they are being intentionally hidden.

This method can be used in modelling insurance losses in general insurance. Loss reporting can influence the loss distributions due to the presence of excess and deductibles on insurance contracts. An insurer can set up excess that may be too low and this can result in more losses being reported, including those could have been covered by the excess. Hence a solution can be a method that does not put fixed excesses. By modifying traditional loss severity distributions used to model operational risk losses, an institution can implement the proposed model to obtain capital charge with missing data. Also an institution can use this method to assess the level at which the datasets are missing information.

7.2 Limitations of study

For a good modelling exercise, there is need for good quality data in terms of granularity, size and consistency. Fitting this model of a dataset that is consistently recorded for different loss types and different business classes would have been ideal. The dataset used for this research was based on publicly recorded loss for different institutions in different countries.

The definition of a single loss was constrained for this research. Several losses can be recorded separately but with the same trigger event and in the same way, a single loss may be recorded when it has multiple triggers. This will significantly influence the loss sizes and frequency and in turn influence the distributions.

© University of Pretoria

The value of β describes the company's loss amount distribution. Converting it to the GR b-value, we get a measure that gives the ratio of large losses to small losses. This measure is used in seismology to assess the level of seismicity in a specific area. Further research needs to be done to assess the level of the b-value that is appropriate for a specific to a financial sector, business line or loss type.

© University of Pretoria

Chapter 8

Conclusion

Modelling operational risk, mainly fraud, is an important risk management exercise. Even though capital requirements are moving towards standardised formulae based methods, there is a need to understand loss distributions for internal management. Understanding the distribution of smaller losses can significantly improve the loss analysis of a company. Problem areas can be identified, for instance, minimum reporting standards can be adjusted to reflect the company's risk exposure.

A method was proposed and developed by Kijko and Smit [17] to model the frequency - magnitude distribution of earthquakes and this method, was modified to model operational risk losses. The loss distribution developed provided an good data fit to the simulated datasets with missing data points. The distribution tapered at the lower tail to represent only the losses that were observed and included in the dataset.

Maximum Likelihood and Method of Moments estimation methods were carried out to get parameter estimates. The estimates were comparative for both methods and with the MLE estimates being more stable and having lower volatility compared to the MME estimates. The most interesting result was how the estimates for β, under

© University of Pretoria

different levels of data deletion, were stable and close to the true parameter value. By including a level of completeness, the β estimates will significantly be different from the true underlying values compared to the estimate when the level of completeness is implicitly defined as done in this research.

8.1 Further Studies

The theory of this proposed method relies on a distribution being able to be transformed to an exponential distribution. Not all distributions can be easily transformed to an exponential distribution which is then coupled with the quadratic function used to define the detection probability distribution. Investigation of other fat-tailed distributions that are usually used to model operational risk losses can be carried out and other functions such as exponential functions as detection function. This would be ideal if the resultant distribution can be represented in closed form.

For this study, simulations were carried out where data were deleted. This makes the data set more biased towards the larger losses. This means that there is need to assess how the proposed approach in this research impacts the upper end of the distribution.

The two methods of estimation used for this research are MME and MLE. This list is not exhaustive and other estimation methods can be used and may produce more stable parameter estimates.

© University of Pretoria

Bibliography

[1] A. D. Brown, "Making sense of the collapse of Barings Bank," *Human Relations*, vol. 58, no. 12, pp. 1579–1604, 2005.

[2] R. J. Rhee, "The Madoff scandal, market regulatory failure and the business education of lawyers," *J. Corp. L.*, vol. 35, p. 363, 2009.

[3] Z. Bodur, "Operational Risk and Operational Risk Related Banking Scandals: Large Incidents," *Maliye Finans Yazıları*, vol. 1, no. 97, 2012.

[4] E. de Jongh, R. de Jongh, D. de Jongh, and G. van Vuuren, "A review of operational risk in banks and its role in the financial crisis," *South African Journal of Economic and Management Sciences*, vol. 16, no. 4, pp. 364–382, 2013, ISSN: 2222-3436. DOI: 10.4102/sajems.v16i4.440. [Online]. Available: http://www.sajems.org/index.php/sajems/article/view/440.

[5] L. Andersen, D. Häger, S. Maberg, M. Næss, and M. Tungland, "The financial crisis in an operational risk management context—a review of causes and influencing factors," *Reliability Engineering and System Safety*, vol. 105, no. Supplement C, pp. 3 –12, 2012, ESREL 2010, ISSN: 0951-8320. DOI: https://doi.org/10.1016/j.ress.2011.09.005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0951832011001773.

© University of Pretoria

[6] Risk.net staff, "Top 10 Operational risks for 2018," *Risk.net*, Feb. 2017, Accessed: 2018-05-27. [Online]. Available: https://www.risk.net/risk-management/5424761/top-10-operational-risks-for-2018.

[7] National Treasury: Republic of South Africa, "Twin Peaks in South Africa: Response And Explanatory Document," National Treasury: Republic of South Africa, Tech. Rep., Dec. 2014.

[8] C. Van Heerden and G. Van Niekerk, "Twin Peaks in South Africa: A new role for the central bank," *Law and Financial Markets Review*, vol. 11, no. 4, pp. 154–162, 2017.

[9] A. Viljoen, V. Lalloo, and L. Hayward, "How should South African financial services firms prepare for change?," 2016, Accessed: 2018-08-27. [Online]. Available: https://www.ey.com/za/en/industries/financial-services/ey-twin-peaks-regulation-in-south-africa.

[10] Solvency Assessment and Management: Pillar I - Sub Committee Capital Requirements Task Group, "Discussion Document 61 (v 4) SCR Standard Formula: Operational Risk," Jul. 2013, Accessed: 09-06-2017. [Online]. Available: www.fsb.co.za/departments/insurance/documents/SAMDiscussion.

[11] A. Chernobai, S. Rachev, and F. Fabozzi, *Operational Risk: A Guide to Basel II Capital Requirements, Models and Analysis*. John Wiley and Sons Inc., 2007.

[12] Basel Committee on Banking Supervision, *Operational risk - supervisory guidelines for the advanced measurement approaches*, Accessed: 17-03-2017, Jun. 2011. [Online]. Available: http://www.bis.org/publ/bcbs196.pdf.

[13] J. Woessner, "Probability of Detecting an Earthquake," *Bulletin of the Seismological Society of America*, vol. 98, pp. 2103–2117, 5 Oct. 2008.

© University of Pretoria

[14] A. Chernobai, C. Menn, M. Moscadelli, S. Rachev, and S. Trueck, "Treatment of incomplete data in the field of operational risk: The effects on parameter estimates, el and ul figures," in *The Advanced Measurement Approach to Operational Risk*, Spain: Risk Books, 2006, pp. 145–168. [Online]. Available: https://eprints.qut.edu.au/21448/.

[15] I. Rozenfield, "Estimating Operational Risk Capital from Truncated Data with Constrained Optimization," Jul. 2012, Accessed: 10-06-2017. [Online]. Available: https://ssrn.com/abstract=2539900.

[16] X. Luo, P. Shevchenko, and J. Donnelly, "Addressing the impact of data truncation and parameter uncertainty on operational risk estimates," *Journal of Operational Risk*, vol. 2, pp. 3–26, 4 Mar. 2006, ISSN: 1755-2710. DOI: 10.21314/JOP.2007.034.

[17] A. Kijko and A. Smit, "Estimation of the Frequency-Magnitude Gutenberg-Richter *b*-value without Making Assumptions on Levels of Completeness," *Seismological Research Letters*, vol. 88, pp. 311–318, 2A Mar. 2017.

[18] H. Berkman and M. E. Bradbury, "Empirical evidence on the corporate use of derivatives," *Financial management*, pp. 5–13, 1996.

[19] Economic Capital (EC) Subgroup, "Specialty Guide on Economic Capital," Society of Actuaries Risk Management Task Force (RMTF), Tech. Rep., Mar. 2004.

[20] J.-H. Wang, Y.-L. Liao, T.-m. Tsai, and G. Hung, "Technology-based financial frauds in Taiwan: Issues and approaches," vol. 2, pp. 1120–1124, 2006.

[21] A. Sharma and P. K. Panigrahi, "A review of financial accounting fraud detection based on data mining techniques," *arXiv preprint arXiv:1309.3944*, 2013.

[22] E. Ngai, Y. Hu, Y. Wong, Y. Chen, and X. Sun, "The application of data mining techniques in financial fraud detection: A classification framework and an aca-

© University of Pretoria

demic review of literature," *Decision Support Systems*, vol. 50, no. 3, pp. 559–569, 2011.

[23] Legal Information Institute: Cornwell Law School, "White-Collar Crime: An Overview," Jun. 2016, Accessed: 2018-09-11. [Online]. Available: https://www.law.cornell.edu/wex/White-collar_crime.

[24] S. W. Buell, "What is securities fraud?" *Duke Law Journal*, vol. 61, p. 511, 2011.

[25] R. B. Thompson and H. A. Sale, "Securities fraud as corporate governance: Reflections upon federalism," *Vanderbilt Law Review*, vol. 56, p. 859, 2003.

[26] J. Cotterill, "McKinsey, KPMG accused of criminal breaches over South Africa Gupta scandal," *Financial Times*, Jan. 2018. [Online]. Available: https://www.ft.com/content/71c6f115-0c5c-33ed-bc00-812263f39d2f.

[27] L. Donnelly, "KPMG woes deepen after VBS bank scandal," *Mail & Guardian*, Apr. 2018. [Online]. Available: https://mg.co.za/article/2018-04-15-kpmg-woes-deepen-after-vbs-bank-scandal.

[28] J. Rossouw, "Steinhoff scandal points to major gaps in stopping unethical corporate behaviour," *The Conversation*, Dec. 2017. [Online]. Available: http://theconversation.com/steinhoff-scandal-points-to-major-gaps-in-stopping-unethical-corporate-behaviour-88905.

[29] A. Lungisa, "The Steinhoff Debacle – the biggest fraud in SA history," *Daily Maverick*, Dec. 2017, Accessed: 2018-05-27. [Online]. Available: https://www.dailymaverick.co.za/opinionista/2017-12-13-the-steinhoff-debacle-the-biggest-fraud-in-sa-history/#.WxpfdNIzbcs.

[30] Basel Committee on Banking Supervision, "International convergence of capital measurement and capital standards," Jun. 2006, Accessed: 17-03-2017. [Online]. Available: http://www.bis.org/publ/bcbs128.pdf.

© University of Pretoria

[31] Solvency II, "Directive 2009/83/EC of the European Parliament and of the Council of 5 November 2002 concerning life assurance," *Official Journal of European Union*, 2002.

[32] F. Aue and M. Kalkbrener, "LDA at Work," *Journal of Operational risk*, vol. 1, pp. 49–93, 4 Feb. 2007.

[33] A. Frachot, P. Georges, and T. Roncalli, "Loss distribution approach for operational risk," Mar. 2001, Accessed: 20-03-2017. [Online]. Available: `https://ssrn.com/abstract=1032523`.

[34] T. Hartung, "Considerations to the quantification of operational risks," University of Munich, Tech. Rep., 2003.

[35] M. G. Cruz, *Modeling, measuring and hedging operational risk*. John Wiley & Sons Chichester, 2002.

[36] M. Moscadelli, "The modelling of operational risk: Experience with the analysis of the data collected by the basel committee," *SSRN Electronic Journal*, Aug. 2004.

[37] P. de Fontnouvelle, V. DeJesus-Rueff, J. Jodan, and E. Rosengren, "Using Loss Data to quantify Operational risk," *SSRN Electronic Journal*, Apr. 2003. DOI: `10.2139/ssrn.395083`. [Online]. Available: `http://dx.doi.org/10.2139/ssrn.395083`.

[38] P. De Fontnouvelle, E. Rosengren, and J. Jordan, "Implications of alternative operational risk modeling techniques," in *The Risks of Financial Institutions*, University of Chicago Press, 2007, pp. 475–512.

[39] K. Dutta and J. Perry, "A Tale of Tails: An Empirical Analysis of Loss Distribution Models for Estimating Operational Risk Capital," Federal Reserve Bank of Boston, Tech. Rep., 2007, Accessed: 2018-05-14. [Online]. Available: `https://ssrn.com/abstract=918880`.

© University of Pretoria

[40] J. B. McDonald and Y. J. Xu, "A generalization of the beta distribution with applications," *Journal of Econometrics*, vol. 66, no. 1-2, pp. 133–152, 1995.

[41] C. Field and M. G. Genton, "The multivariate g-and-h distribution," *Technometrics*, vol. 48, no. 1, pp. 104–111, 2006.

[42] G. E. Willmot, X. S. Lin, and X. S. Lin, *Lundberg approximations for compound distributions with insurance applications*. Springer Science & Business Media, 2001, vol. 156.

[43] E. Peköz and S. M. Ross, "Compound random variables," *Probability in the Engineering and Informational Sciences*, vol. 18, no. 4, pp. 473–484, 2004.

[44] R. A. Fisher and L. H. C. Tippett, "Limiting forms of the frequency distribution of the largest or smallest member of a sample," in *Mathematical Proceedings of the Cambridge Philosophical Society*, Cambridge University Press, vol. 24, 1928, pp. 180–190.

[45] R.-D. Reiss, M. Thomas, and R. Reiss, "Statistical analysis of extreme values," vol. 2, 2007.

[46] A. J. McNeil, "Extreme value theory for risk managers," *Departement Mathematik ETH Zentrum*, 1999.

[47] A. J. McNeil and R. Frey, "Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach," *Journal of empirical finance*, vol. 7, no. 3, pp. 271–300, 2000.

[48] M. Loretan and P. C. Phillips, "Testing the covariance stationarity of heavy-tailed time series: An overview of the theory with applications to several financial datasets," *Journal of empirical finance*, vol. 1, no. 2, pp. 211–248, 1994.

[49] A. F. Laurens de Haan, "Extreme Value Theory: An Introduction," Springer Series in Operations Research, 2006.

© University of Pretoria

[50] A. Butler. (Mar. 2007). Advantages and disadvantages of extreme value methods, [Online]. Available: `http://www.bioss.ac.uk/people/adam/teaching/OR_EVT/2007/node12.html`.

[51] P. Embrechts, "Extreme Value Theory: Potentials and Limitations as an Integrated Risk Management Tool," *Derivatives use, Trading and Regulation*, vol. 6, Feb. 2000.

[52] C. T. Spathis, "Detecting false financial statements using published data: some evidence from Greece," *Managerial Auditing Journal*, vol. 17, no. 4, pp. 179–191, 2002.

[53] C. J. Skousen and C. J. Wright, "Contemporaneous risk factors and the prediction of financial statement fraud," *SSRN Electronic Journal*, Aug. 2006.

[54] D. J. Hand, "Principles of data mining," *Drug safety*, vol. 30, no. 7, pp. 621–622, 2007.

[55] M Bharati and R., "Data Mining Techniques and Applications," *Indian Journal of Computer Science and Engineering*, vol. 1, pp. 301–305, 4 Dec. 2010.

[56] E. Kirkos, C. Spathis, and Y. Manolopoulos, "Data mining techniques for the detection of fraudulent financial statements," *Expert systems with applications*, vol. 32, no. 4, pp. 995–1003, 2007.

[57] J. P. Jiawei Han Micheline Kamber, "Data Mining: Concepts and Techniques," The Morgan Kaufmann Series in Data Management Systems, 2006.

[58] F. H. Glancy and S. B. Yadav, "A computational model for financial reporting fraud detection," *Decision Support Systems*, vol. 50, no. 3, pp. 595–601, 2011.

[59] A. S. Koyuncugil and N. Ozgulbas, "Financial Profiling for Detecting Operational Risk by Data Mining," *World Academy of Science, Engineering and Technology*, vol. 46, 2008.

© University of Pretoria

[60] E. I. Altman, “Financial ratios, discriminant analysis and the prediction of corporate bankruptcy,” *The journal of finance*, vol. 23, no. 4, pp. 589–609, 1968.

[61] A. Berson, S. Smith, and K. Thearling, *Building data mining applications for CRM*. McGraw-Hill Professional, 1999.

[62] S. G. Eick and D. E. Fyock, “Visualizing corporate data,” *AT&T technical journal*, vol. 75, no. 1, pp. 74–86, 1996.

[63] T. Kohonen, “The self-organizing map,” *Neurocomputing*, vol. 21, no. 1-3, pp. 1–6, 1998.

[64] E. Corchado and Á. Herrero, “Neural visualization of network traffic data for intrusion detection,” *Applied Soft Computing*, vol. 11, no. 2, pp. 2042–2056, 2011.

[65] K. Šimunić, “Visualization of Stock Market Charts,” in *In Proceedings from the 11th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision 2003 (2003), Plzen-Bory (CZ), 2003*, 2003.

[66] M. L. Huang, J. Liang, and Q. V. Nguyen, “A visualization approach for frauds detection in financial market,” in *Information Visualisation, 2009 13th International Conference*, IEEE, 2009, pp. 197–202.

[67] D. Olszewski, “Fraud detection using self-organizing map visualizing the user profiles,” *Knowledge-Based Systems*, vol. 70, pp. 324–334, 2014.

[68] E. N. Argyriou, A. A. Sotiraki, and A. Symvonis, “Occupational fraud detection through visualization,” in *Intelligence and Security Informatics (ISI), 2013 IEEE International Conference on*, IEEE, 2013, pp. 4–6.

[69] D. Yue, X. Wu, Y. Wang, Y. Li, and C.-H. Chu, “A review of data mining-based financial fraud detection research,” in *Wireless Communications, Networking and Mobile Computing, 2007. WiCom 2007. International Conference on*, Ieee, 2007, pp. 5519–5522.

© University of Pretoria

[70] N. Gatzert and H. Wesker, "A comparative assessment of Basel II/III and Solvency II," *The Geneva Papers on Risk and Insurance-Issues and Practice*, vol. 37, no. 3, pp. 539–570, 2012.

[71] G. C. A. EUROPEEN, *Comparison of the Regulatory Approach in Insurance and Banking in the Context of Solvency II*, Accessed: 12-06-2018, 2013. [Online]. Available: https://actuary.eu/documents/SII\%20vs\%20Basel\%20II_Dec_12_final.pdf.

[72] A. Cornford, "Basel II: The Revised Framework Of June 2004," United Nations Conference on Trade and Development, UNCTAD Discussion Papers 178, Apr. 2005. [Online]. Available: http://unctad.org/en/docs/osgdp20052_en.pdf.

[73] L. Balthazar, "From Basel 1 to Basel 3," in *From Basel 1 to Basel 3: The Integration of State-of-the-Art Risk Modeling in Banking Regulation*, Springer, 2006, pp. 209–213.

[74] Basel Committee on Banking Supervision, *Standardised measurement approach for operational risk*, Accessed: 11-05-2017, Mar. 2016. [Online]. Available: https://www.bis.org/bcbs/publ/d355.pdf.

[75] A. Hater and M. Kronbichler, "Capital requirements for operational risk – new SMA," Mar. 2017, Accessed: 2018-05-11. [Online]. Available: www.bankinghub.eu/banking/finance-risk/capital-requirements-operational-risk-new-sma.

[76] P.-E. Chabanel, "Proposed Capital Framework for Operational Risk," Moody's Analytics, Tech. Rep., 2017, Accessed: 2018-05-11. [Online]. Available: https://www.moodysanalytics.com/-/media/whitepaper/2017/proposed-capital-framework-for-operational-risk.pdf.

[77] D. Serbee, A. Heimrich, G. Allbutt, and H. Katz, "Ten key points from Basel's Standardized Measurement Approach for operational risk," PWC, Tech. Rep.,

© University of Pretoria

Mar. 2016, Accessed: 2018-05-11. [Online]. Available: https://www.pwc.com/us/en/financial-services/regulatory-services/publications/assets/basel-operational-risk-proposal-march-2016.pdf.

[78] KPMG, "Basel 4: The way ahead," KPMG, Tech. Rep., Feb. 2018, Accessed: 2018-05-11. [Online]. Available: https://assets.kpmg.com/content/dam/kpmg/be/pdf/2018/03/basel-4-the-way-ahead.pdf.

[79] Solvency II, "Directive 2009/138/EC of the European Parliament and of the Council of 25 November 2009 on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II)," *Official Journal of European Union*, 2009, Accessed: 12-07-2018. [Online]. Available: https://www.tsb.org.tr/images/Documents/SolvencyIIDirektifi.pdf.

[80] Canadian Institute of Actuaries, *Research paper on operational risk.* Accessed: 17-03-2017, Dec. 2014. [Online]. Available: http://www.cia-ica.ca/docs/default-source/2014/214118e.pdf.

[81] European Insurance and Occupational Pensions Authority, *The underlying assumptions in the standard formula for the solvency capital requirement calculation*, Accessed: 12-03-2018, Dec. 2014. [Online]. Available: https://eiopa.europa.eu/Publications/Standards/EIOPA-14-322_Underlying_Assumptions.pdf.

[82] J. Corrigan and P. Luraschi, "Operational risk modelling framework," in *Milliman Research Report*, Milliman, Feb. 2013. [Online]. Available: http://www.milliman.com/uploadedFiles/insight/life-published/operational-risk-modelling-framework.pdf.

[83] T. Mori, E. Harada, *et al.*, "Internal measurement approach to operational risk capital charge," Bank of Japan, Tech. Rep., 2001.

© University of Pretoria

[84] V. Chavez-Demoulin, P. Embrechts, and J. Nešlehová, "Quantitative models for operational risk: Extremes, dependence and aggregation," *Journal of Banking & Finance*, vol. 30, no. 10, pp. 2635–2658, 2006.

[85] P. Embrechts, C. KLüppelberg, and T. Mikosch, *Modelling extremal events: for insurance and finance*. Springer Science & Business Media, 2013, vol. 33.

[86] J. Beirlant, Y. Goegebeur, J. Segers, and J. L. Teugels, *Statistics of extremes: theory and application*. John Wiley & Sons, 2006.

[87] N. Enrique, "Practical calculation of expected and unexpected losses in operational risk by simulation methods," *Banca and Finanzas: Documentos de Trabajo*, vol. 1, pp. 1–12, 1 Oct. 2006.

[88] D. Hendricks, "Evaluation of value-at-risk models using historical data," *Economic Policy Review*, vol. 2, no. 1, 1997.

[89] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath, "Coherent measures of risk," *Mathematical finance*, vol. 9, no. 3, pp. 203–228, 1999.

[90] S. Coles, J. Bawa, L. Trenner, and P. Dorazio, *An introduction to statistical modeling of extreme values*. Springer, 2001, vol. 208.

[91] J. Beirlant, Y. Goegebeur, J. Segers, and J. L. Teugels, *Statistics of extremes: theory and applications*. John Wiley & Sons, 2006.

[92] P. Thompson, Y. Cai, D. Reeve, and J. Stander, "Automated threshold selection methods for extreme wave analysis," *Coastal Engineering*, vol. 56, no. 10, pp. 1013–1021, 2009.

[93] B. Gutenberg and C. Richter, "Magnitude and Energy of Earthquakes Part 1," *Bulletin of the Seismological Society of America*, vol. 32, pp. 163–191, 3 Jul. 1942.

[94] B. Gutenberg and C. Richter, "Magnitude and Energy of Earthquakes Part 2," *Bulletin of the Seismological Society of America*, vol. 46, pp. 105–145, 2 Apr. 1956.

© University of Pretoria

[95] S. Wiemer and M. Wyss, “Minimum magnitude of completeness in earthquake catalogs: Examples from alaska, the western united states, and japan,” *Bulletin of the Seismological Society of America*, vol. 90, pp. 859–869, 4 Aug. 2000.

[96] D. Amorese, “Applying a Change - Point Detection Method on Frequency - Magnitude Distributions,” *The Bulletin of the Seismological Society of America*, vol. 97, pp. 1742–1749, Oct. 2007. DOI: 10.1785/0120060181.

[97] A. Cao and S. S. Gao, “Temporal variation of seismic *b*-values beneath northeastern Japan island arc,” *Geophysical Research Letters*, vol. 29, no. 9, pp. 48–1–48–3, 2002, ISSN: 1944-8007. DOI: 10.1029/2001GL013775. [Online]. Available: http://dx.doi.org/10.1029/2001GL013775.

[98] Y. Ogata and K. Katsura, “Analysis of temporal and spatial heterogeneity of magnitude frequency distribution inferred from earthquake catalogues,” *Geophysical Journal International*, vol. 113, pp. 727–738, 3 Jun. 1993.

[99] J. Woessner and S. Wiemer, “Assessing the Quality of Earthquake Catalogues: Estimating the Magnitude of Completeness and Its Uncertainty,” *Bulletin of the Seismological Society of America*, vol. 95, pp. 684–698, 2 Apr. 2005.

[100] A. Mignan, “Functional shape of the earthquake frequency-magnitude distribution and completeness magnitude,” *Journal of Geophysical Research*, vol. 117, B8 Aug. 2012.

[101] J. L. Devore and K. N. Berk, *Modern Mathematical Statistics with Applications*. Springer Science+Business Media, 2012.

[102] S. I. e. Abramowitz M., *Handbook of mathematical functions (without numerical tables)*, 10th ed. NBS, 1972, ISBN: 505-506-509-5.

[103] N. B. Norman L. Johnson Samuel Kotz, *Continuous Univariate Distributions, Vol. 1 (Wiley Series in Probability and Statistics)*, 2nd ed. Wiley-Interscience, 1994, vol. 1, ISBN: 0471584959,9780471584957.

© University of Pretoria

Appendix I

Mean and Variance

Loss Frequency Distributions		
Distribution	**Mean of distribution**	**Variance of distribution**
Binomial	np	np(1-p)
Geometric	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Negative Binomial [1]	nβ	$n\beta(1+\beta)$
Poisson	λ	λ

Table 8.1: Formulae for mean and variance of loss frequency distribution

$$\Gamma(\alpha; \beta x) = \int_0^{\beta x} t^{\alpha-1} e^{-t} dt \tag{8.1}$$

$$I(x; \alpha, \beta) = \frac{\int_0^x t^{\alpha-1}(1-t)^{\beta-1} dt}{\int_0^1 t^{\alpha-1}(1-t)^{\beta-1} dt} = \frac{B(x; \alpha, \beta)}{B(\alpha, \beta)} \tag{8.2}$$

© University of Pretoria

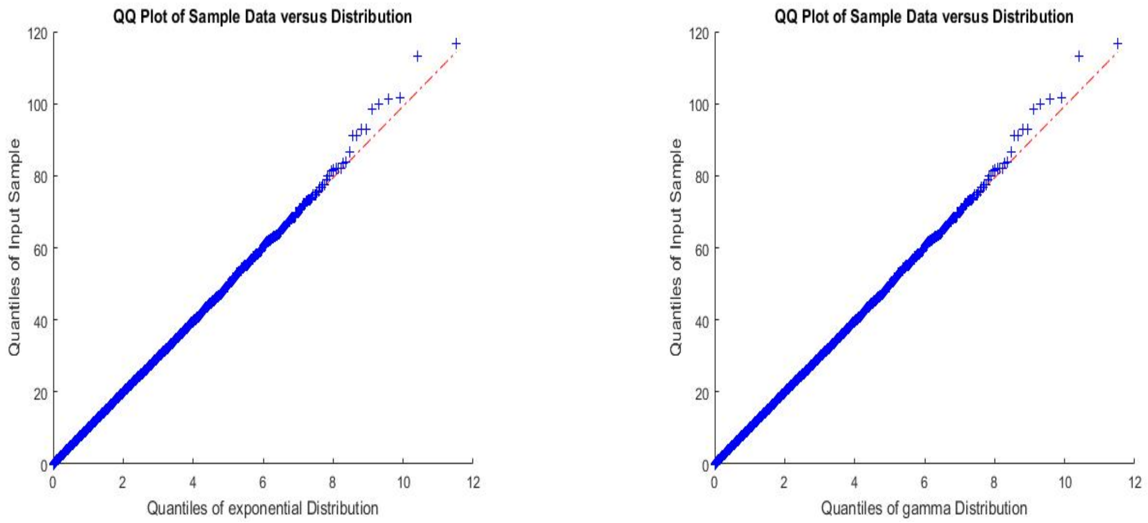
Loss Amount Distributions		
Distribution	**Mean of distribution**	**Variance of distribution**
Exponential	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Lognormal	$e^{\mu+\frac{\sigma^2}{2}}$	$\left(e^{\sigma^2}-1\right)e^{2\mu+\sigma^2}$
Weibull	$\beta^{-\frac{1}{\alpha}}\Gamma\left(1+\frac{1}{\alpha}\right)$	$\beta^{-\frac{2}{\alpha}}\left(\Gamma\left(1+\frac{2}{\alpha}\right)-\Gamma^2\left(1+\frac{1}{\alpha}\right)\right)$
Gamma	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$
Beta	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha=\beta)^2(\alpha+\beta+1)}$
Pareto	$\frac{\alpha\beta}{\alpha-1}$	$\frac{\alpha\beta^2}{(\alpha-1)^2(\alpha-2)}$
Burr	$\frac{\beta^{\frac{1}{\gamma}}}{\Gamma(\alpha)}\Gamma\left(1+\frac{1}{\gamma}\right)\Gamma\left(\alpha-\frac{1}{\gamma}\right)^2$	$\frac{\beta^{\frac{2}{\gamma}}}{\Gamma(\alpha)}\Gamma\left(1+\frac{2}{\gamma}\right)\Gamma\left(\alpha-\frac{2}{\gamma}\right)-\frac{\beta^{\frac{2}{\gamma}}}{\Gamma^2(\alpha)}\Gamma^2\left(1+\frac{1}{\gamma}\right)\Gamma^2\left(\alpha-\frac{1}{\gamma}\right)^3$

Table 8.2: Formulae for mean and variance of loss amount distribution

Simulated Dataset

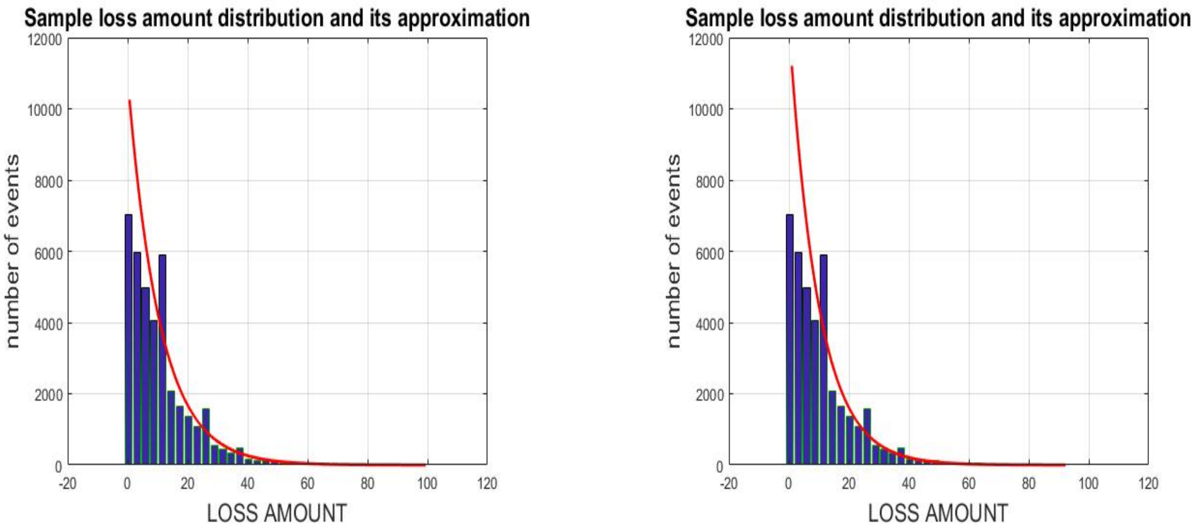
Below are the graphs for the exponentially distributed and the QQ-plots to assess how the level of deletion have on the distribution of the losses.

© University of Pretoria



(a) *(b)*

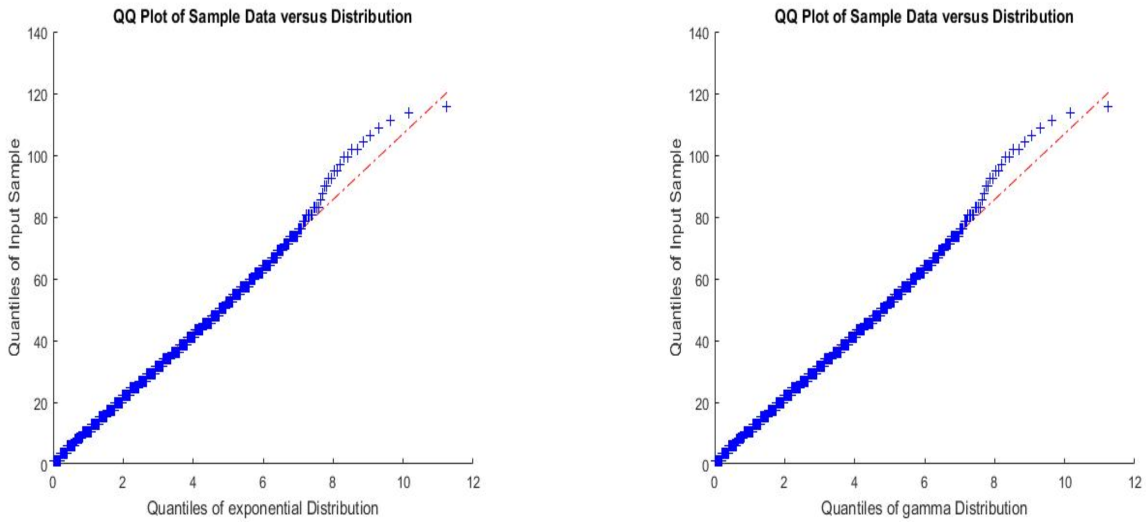
Figure 8.1: QQ plots for $\alpha = 1$



(a) Method of moments *(b) Maximum likelihood*

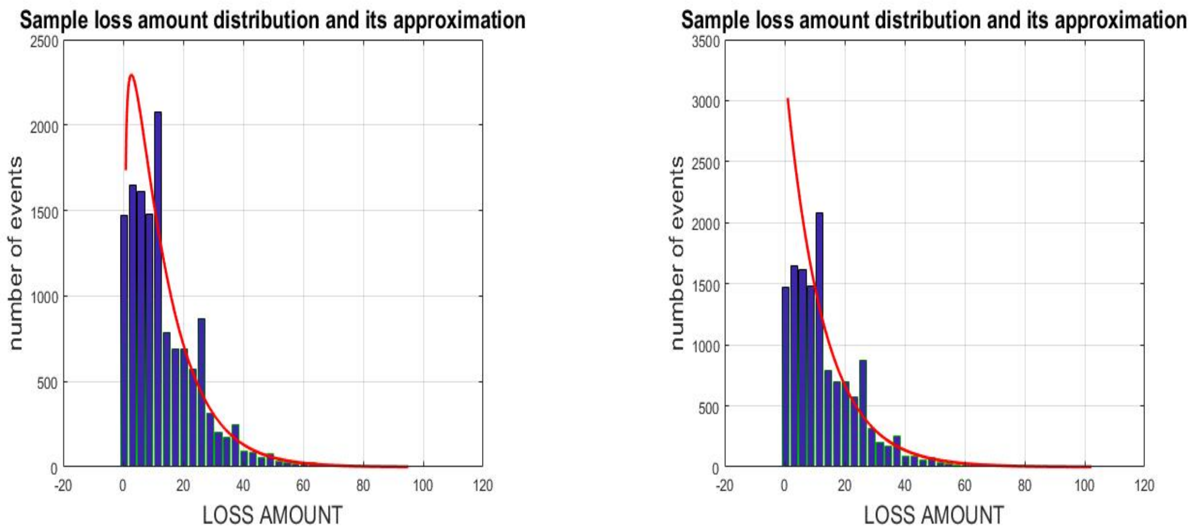
Figure 8.2: Number of losses for incomplete dataset 1.1

© University of Pretoria



(a) *(b)*

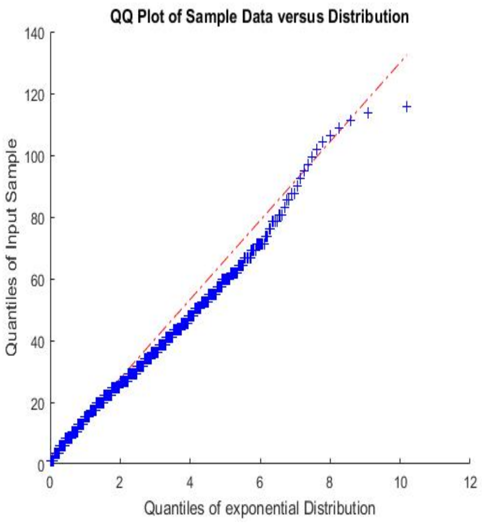
Figure 8.3: QQ plots for $\alpha = 1.1$



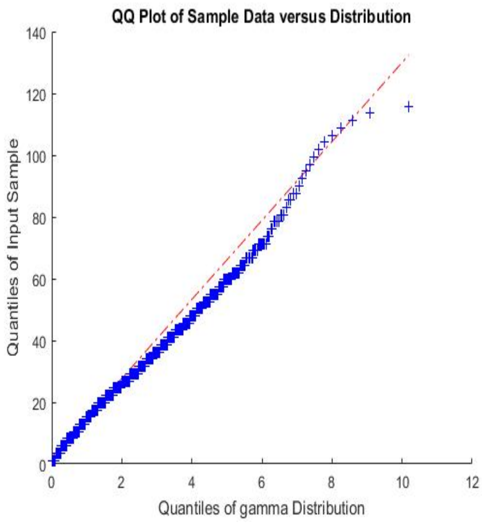
(a) Method of moments *(b) Maximum likelihood*

Figure 8.4: Number of losses for incomplete dataset 1.5

© University of Pretoria

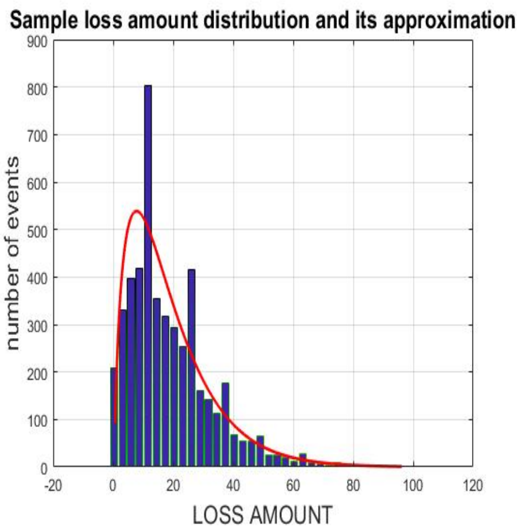


(a)

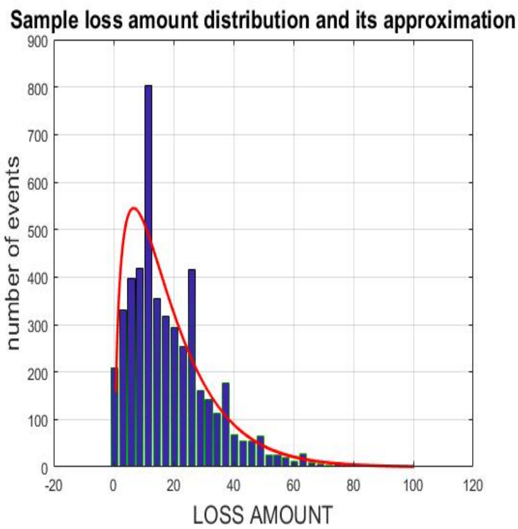


(b)

Figure 8.5: QQ plots for $\alpha = 1.5$



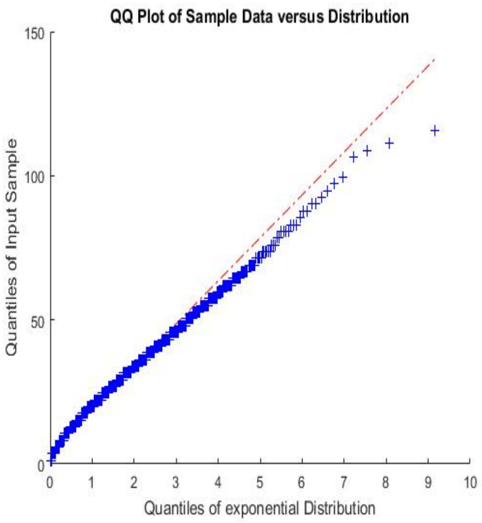
(a) Method of moments



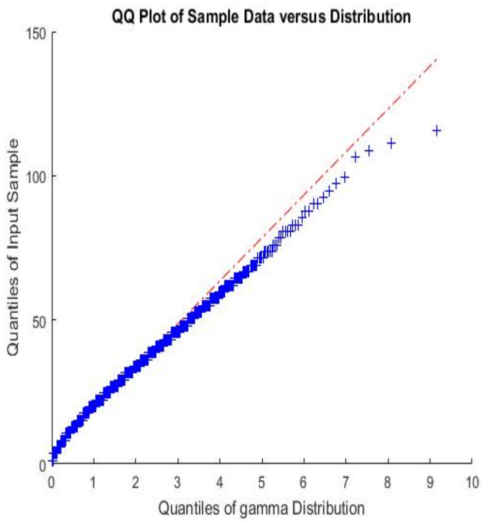
(b) Maximum likelihood

Figure 8.6: Number of losses for incomplete dataset 2

© University of Pretoria

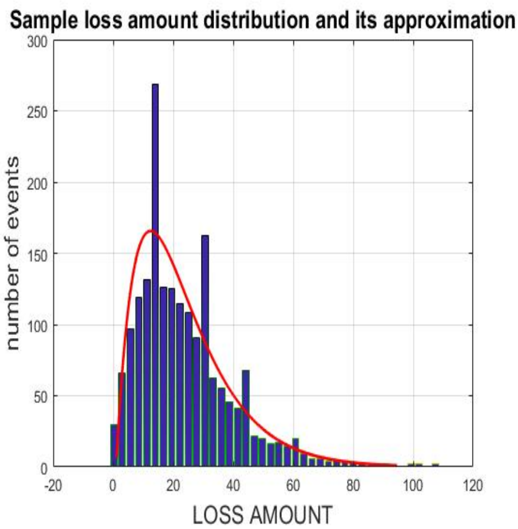


(a)

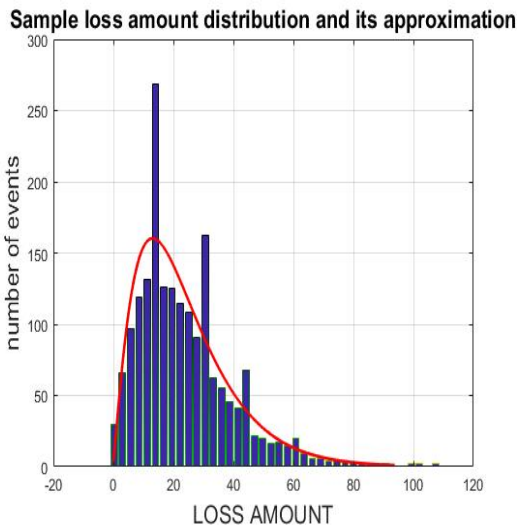


(b)

Figure 8.7: QQ plots for $\alpha = 2$



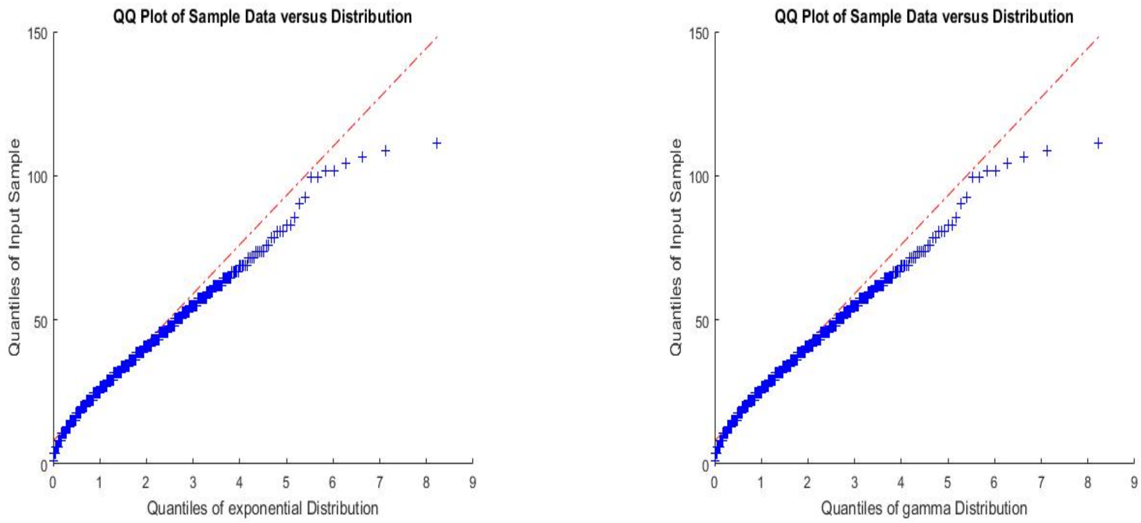
(a) Method of moments



(b) Maximum likelihood

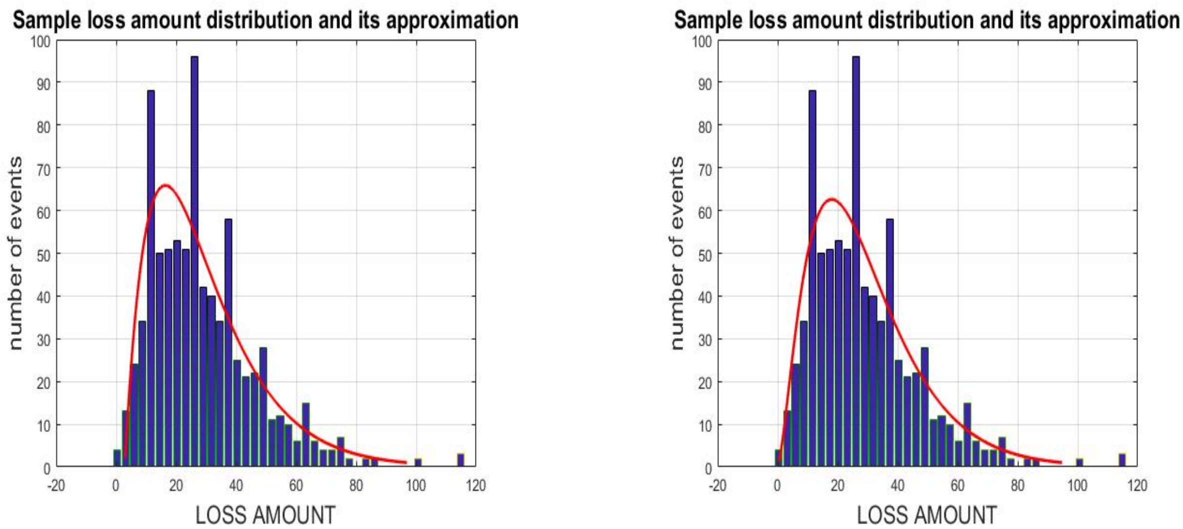
Figure 8.8: Number of losses for incomplete dataset 2.5

© University of Pretoria



(a) *(b)*

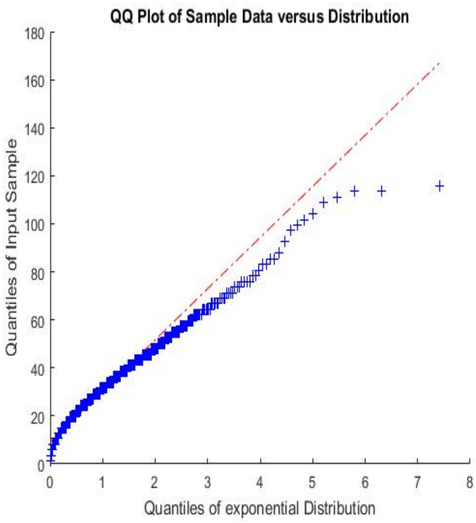
Figure 8.9: QQ plots for $\alpha = 2.5$



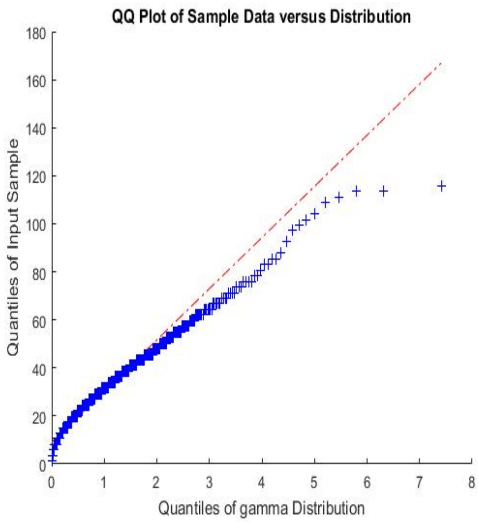
(a) Method of moments *(b) Maximum likelihood*

Figure 8.10: Number of losses for incomplete dataset 3

© University of Pretoria

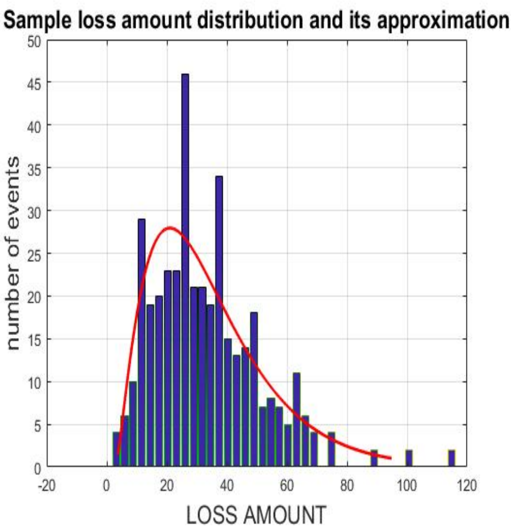


(a)

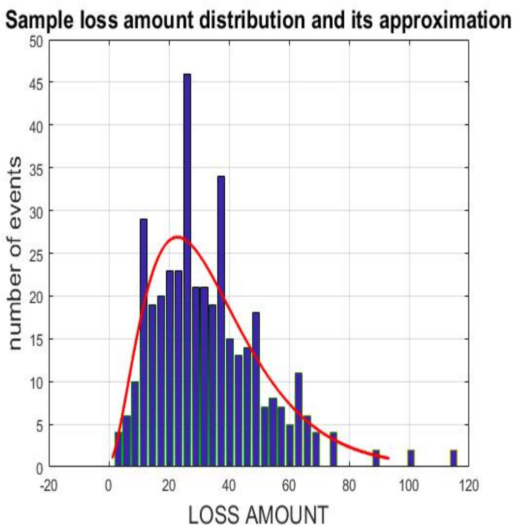


(b)

Figure 8.11: QQ plots for $\alpha = 3$



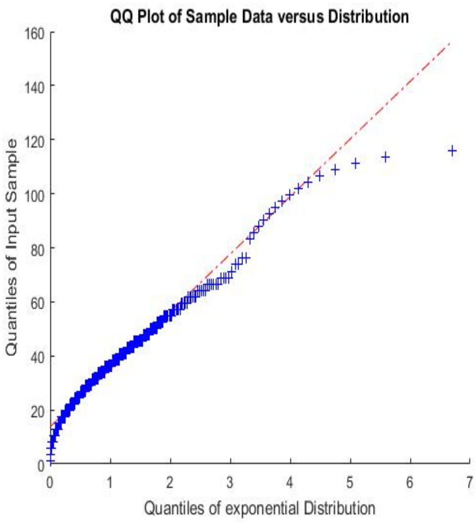
(a) Method of moments



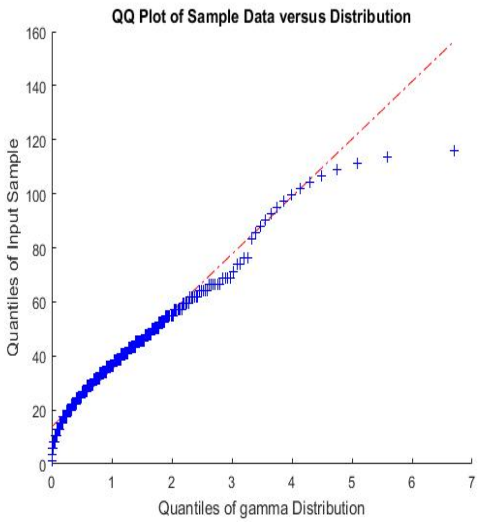
(b) Maximum likelihood

Figure 8.12: Number of losses for incomplete dataset 3.5

© University of Pretoria

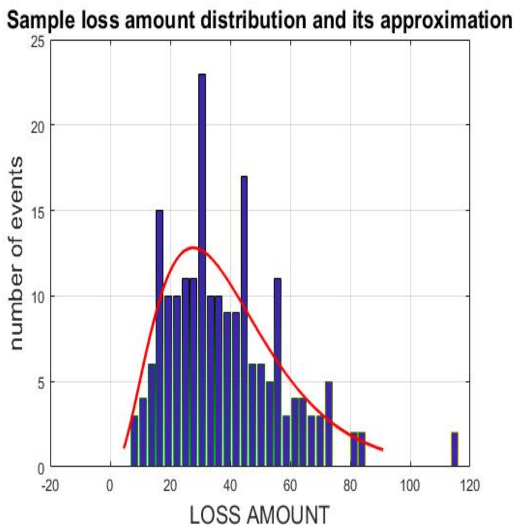


(a)

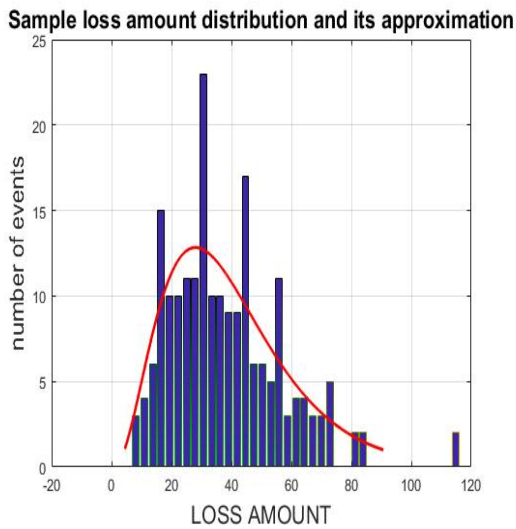


(b)

Figure 8.13: QQ plots for $\alpha = 3.5$



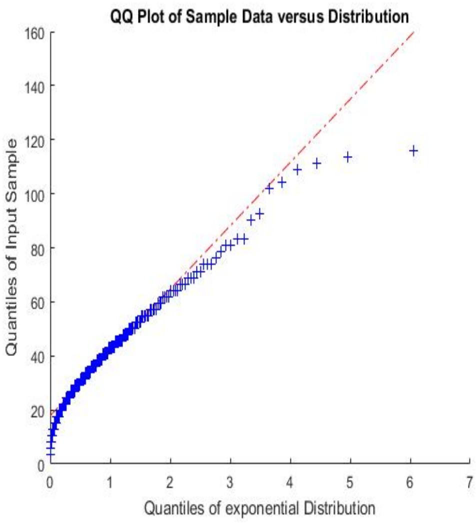
(a) Method of moments



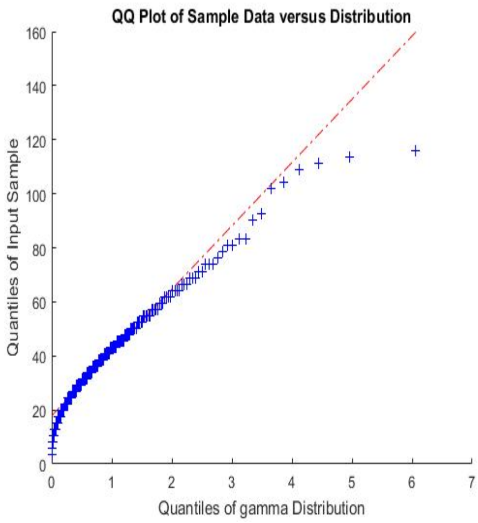
(b) Maximum likelihood

Figure 8.14: Number of losses for incomplete dataset 4

© University of Pretoria

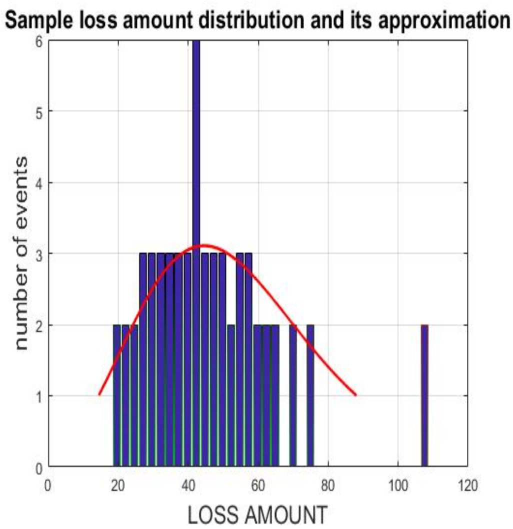


(a)

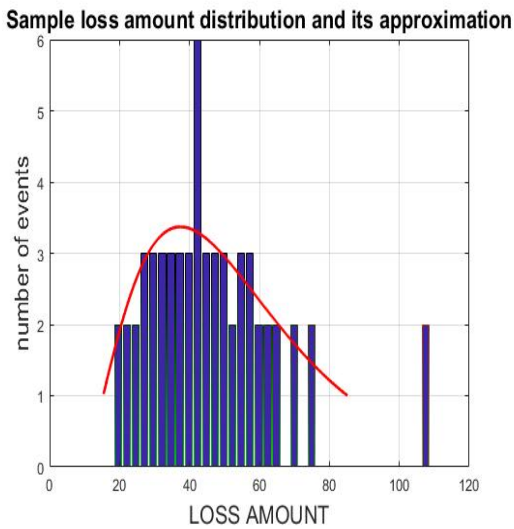


(b)

Figure 8.15: QQ plots for $\alpha = 4$



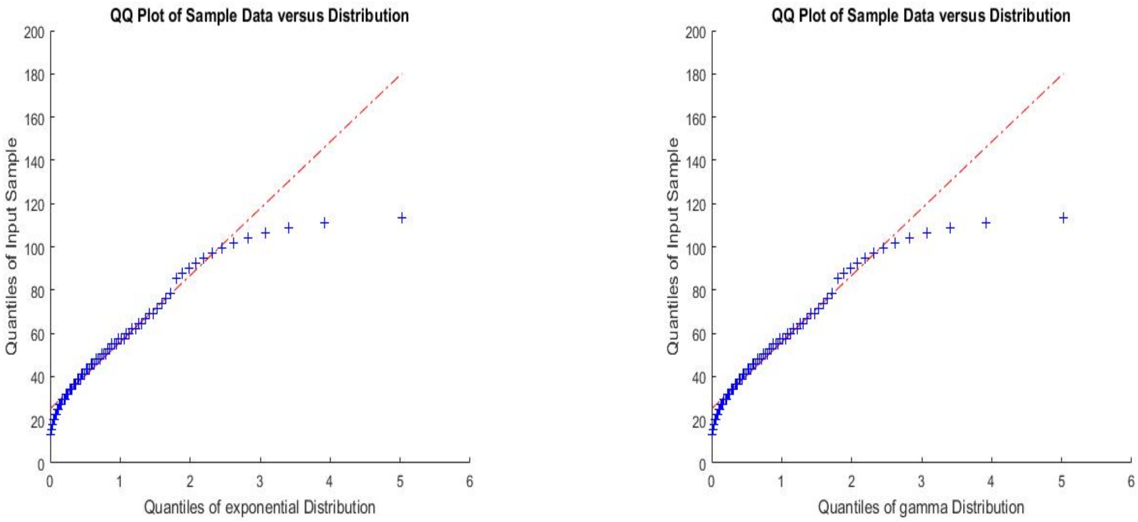
(a) Method of moments



(b) Maximum likelihood

Figure 8.16: Number of losses for incomplete dataset 5

© University of Pretoria



(a) *(b)*

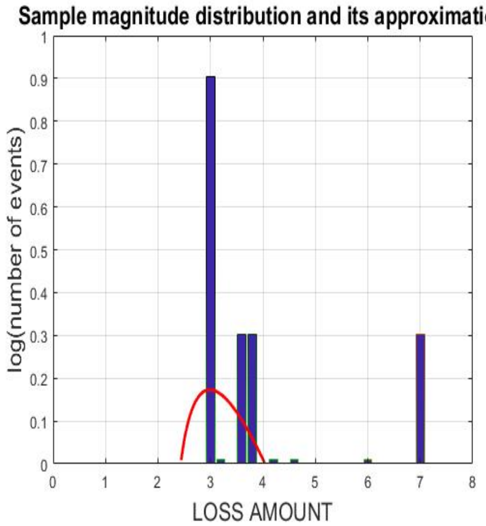
Figure 8.17: QQ plots for $\alpha = 5$

© University of Pretoria

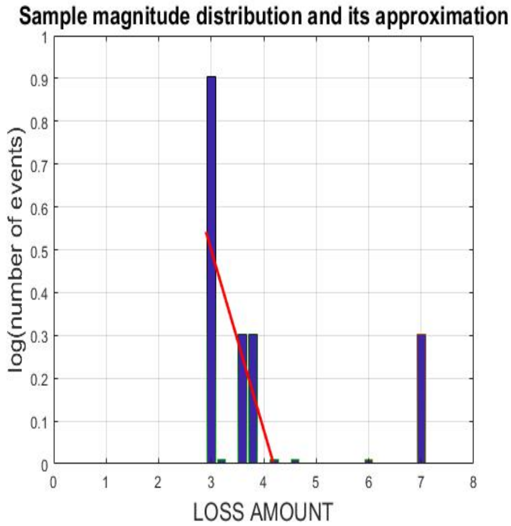
WillisRE dataset

The loss data collected by Willis RE was split according to business line and the loss type. Loss distributions were plotted for each loss type and for each business line. Below are the frequency distributions

Data split according to Business Lines

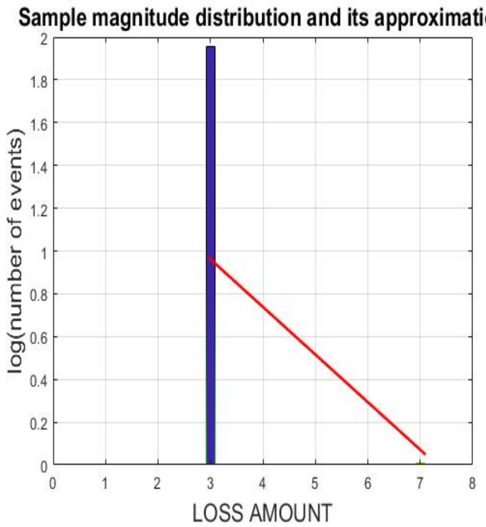


(a) *Method of moments*

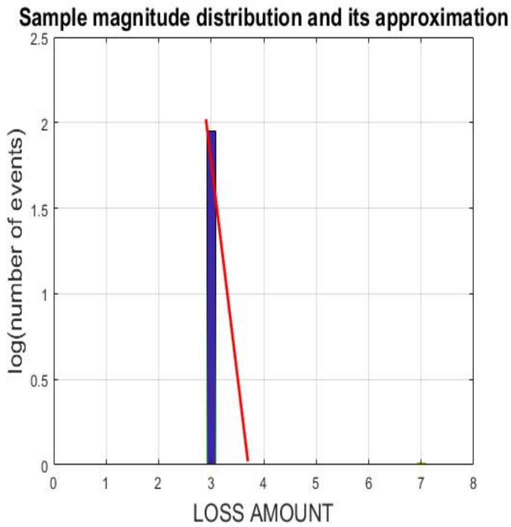


(b) *Maximum likelihood*

Figure 8.18: Business Line: Agency Services



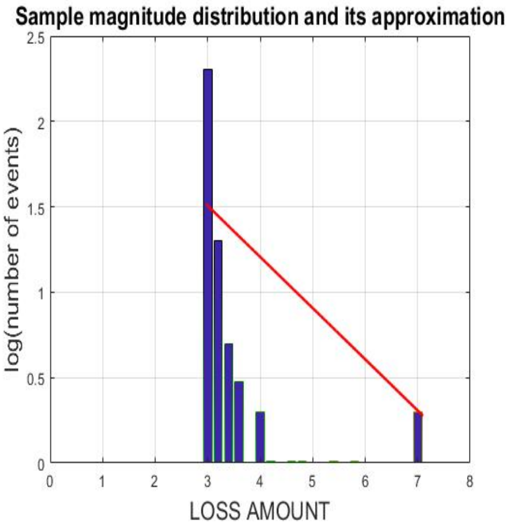
(a) *Method of moments*



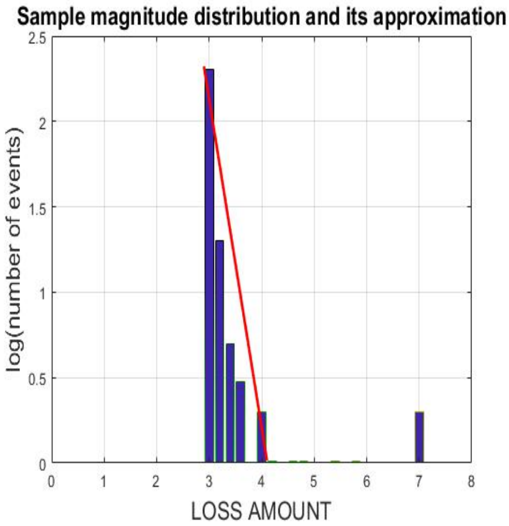
(b) *Maximum likelihood*

Figure 8.19: Business Line: Asset Management

© University of Pretoria

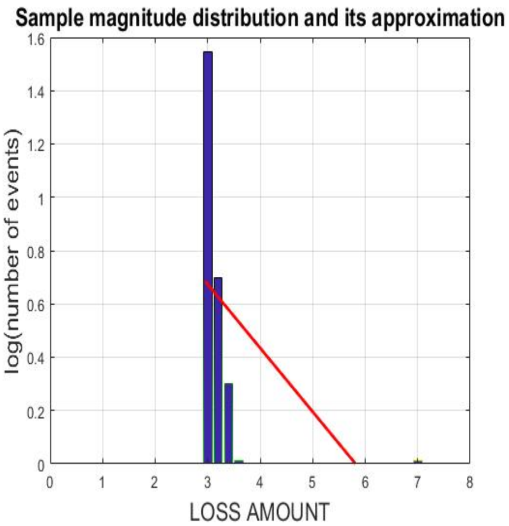


(a) *Method of moments*

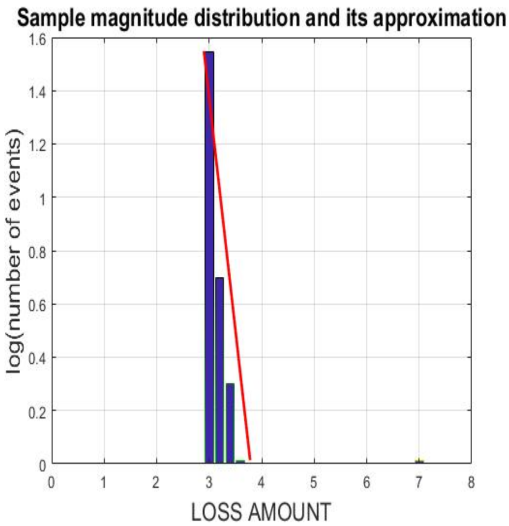


(b) *Maximum likelihood*

Figure 8.20: Business Line: Commercial Banking



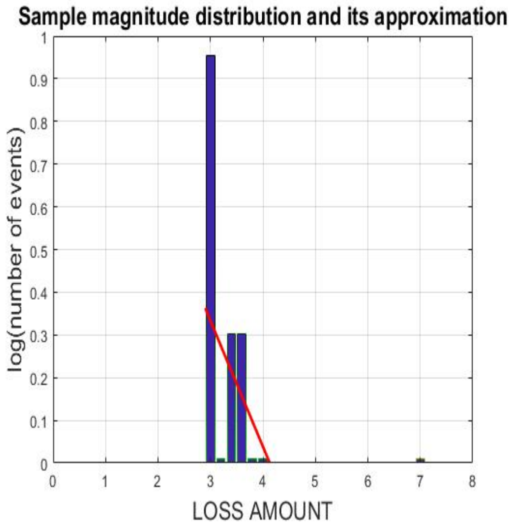
(a) *Method of moments*



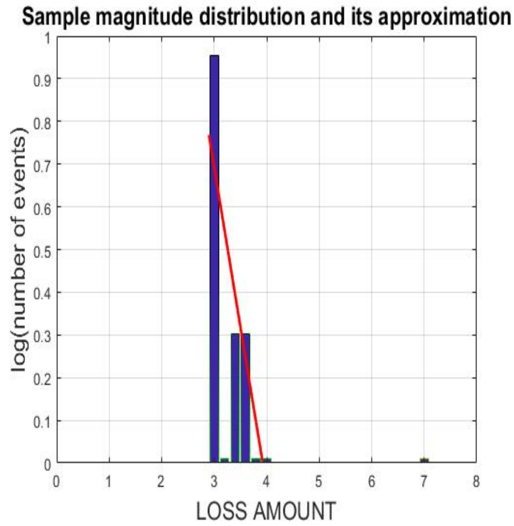
(b) *Maximum likelihood*

Figure 8.21: Business Line: Corporate Finance

© University of Pretoria

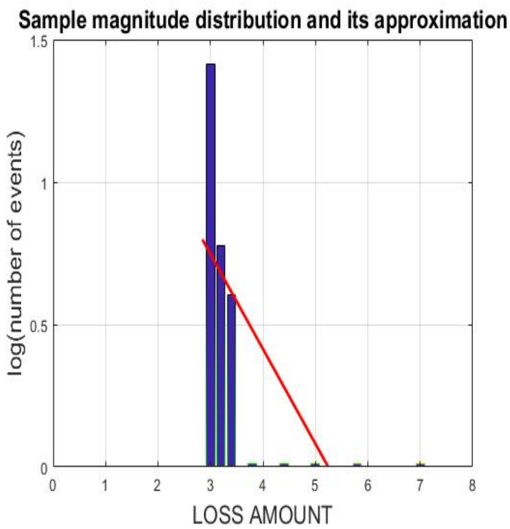


(a) Method of moments

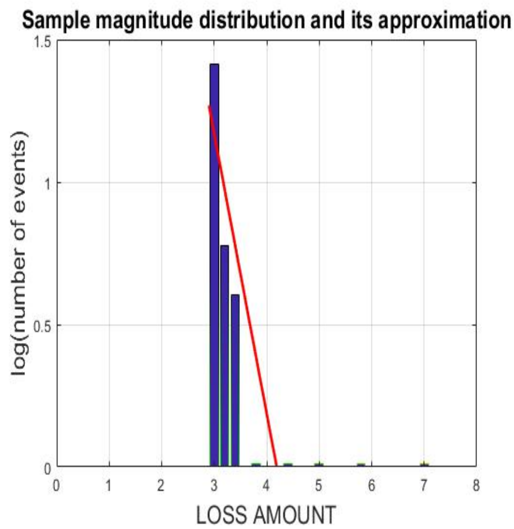


(b) Maximum likelihood

Figure 8.22: Business Line: Life Insurance



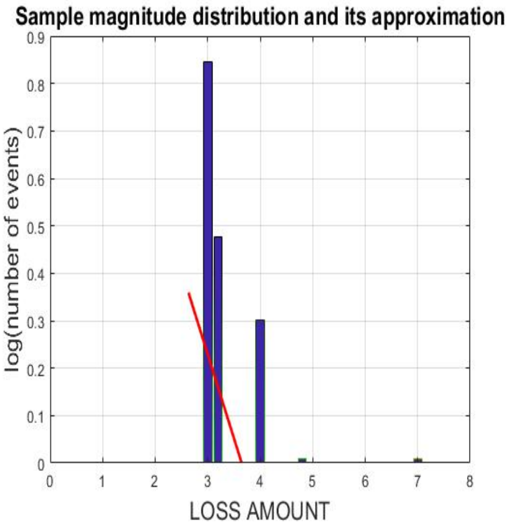
(a) Method of moments



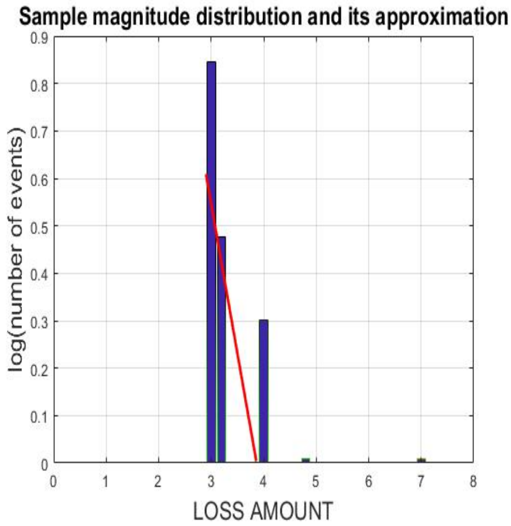
(b) Maximum likelihood

Figure 8.23: Business Line: Non-life Insurance

© University of Pretoria

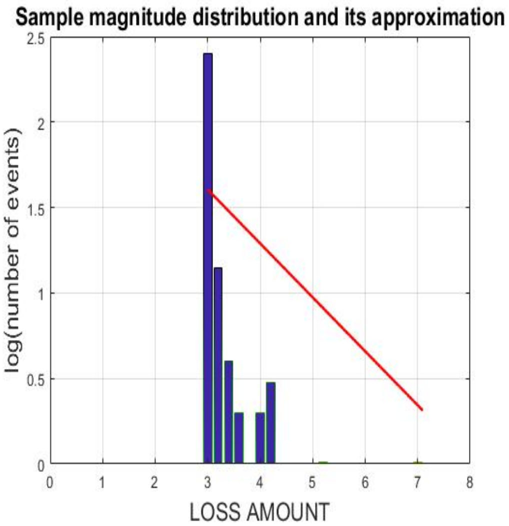


(a) Method of moments

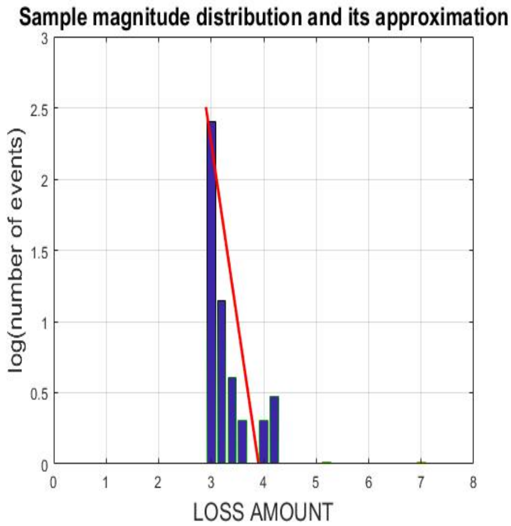


(b) Maximum likelihood

Figure 8.24: Business Line: Payment and Settlement



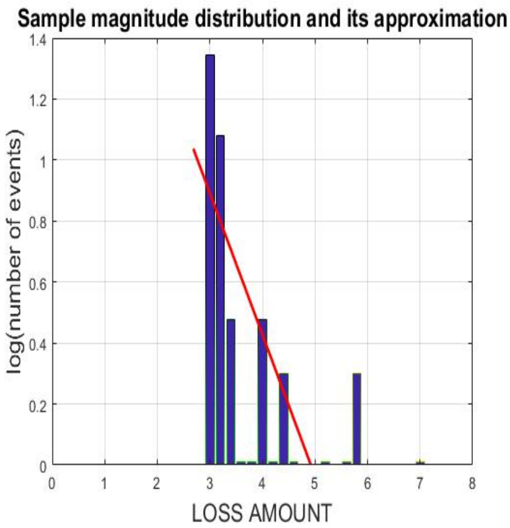
(a) Method of moments



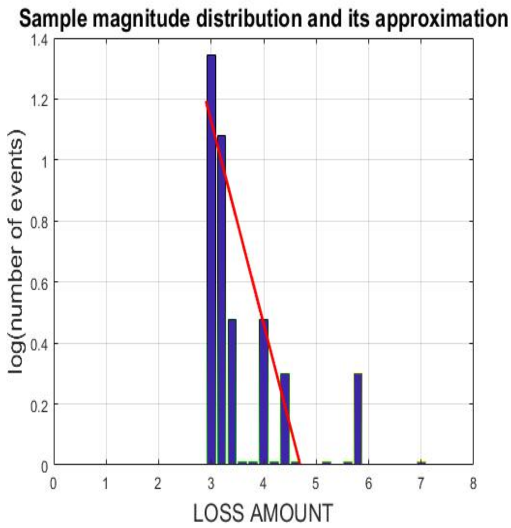
(b) Maximum likelihood

Figure 8.25: Business Line: Retail Banking

© University of Pretoria

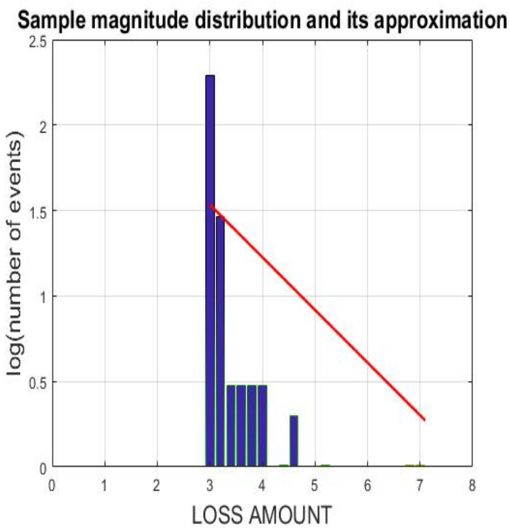


(a) *Method of moments*

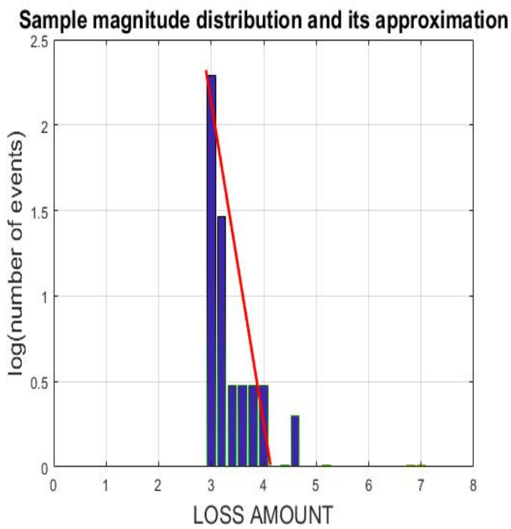


(b) *Maximum likelihood*

Figure 8.26: Business Line: Retail Brokerage



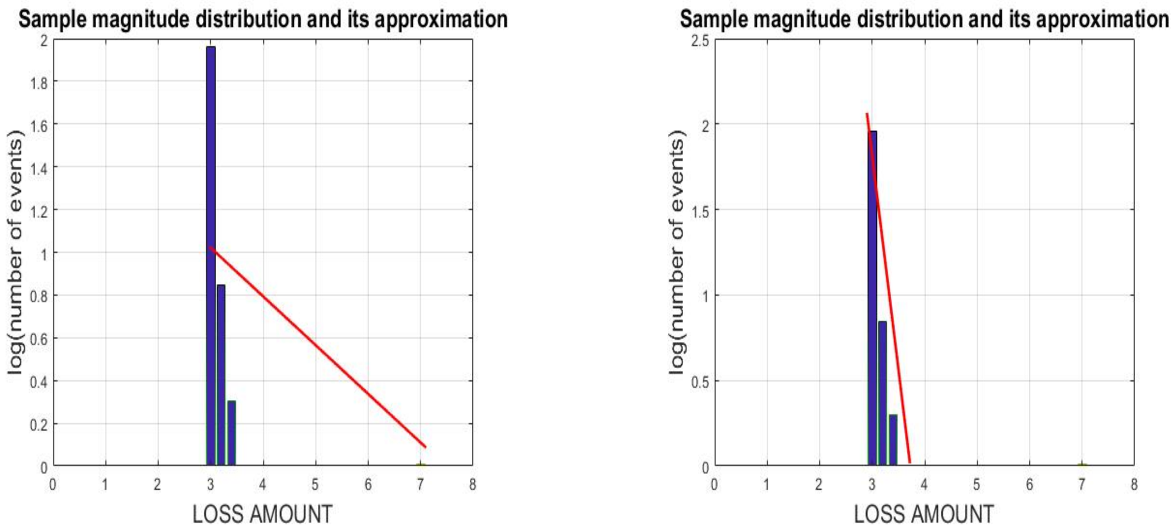
(a) *Method of moments*



(b) *Maximum likelihood*

Figure 8.27: Business Line: Trading and Sales

© University of Pretoria

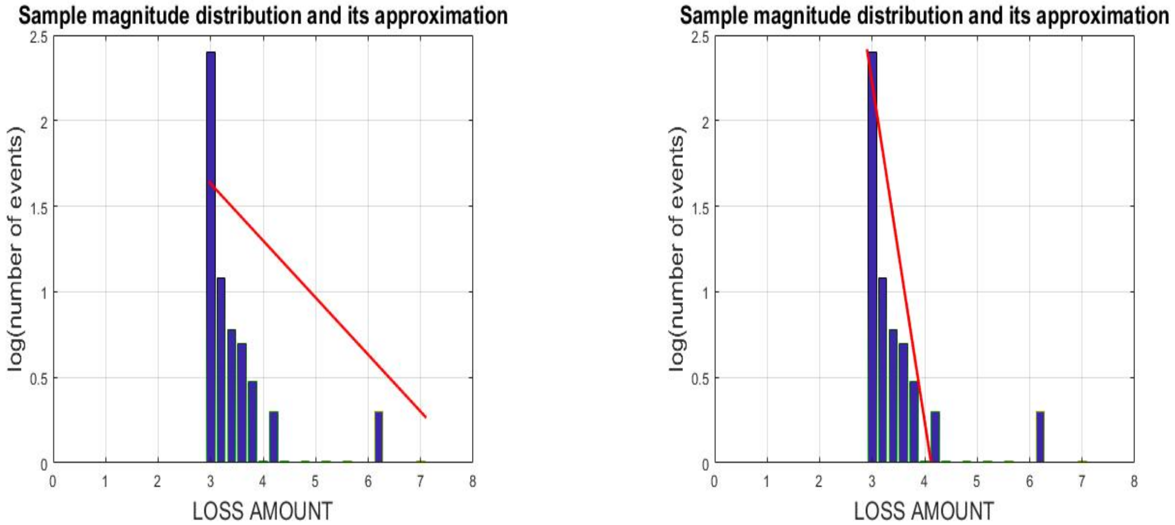


(a) Method of moments *(b) Maximum likelihood*

Figure 8.28: Business Line: Unspecified

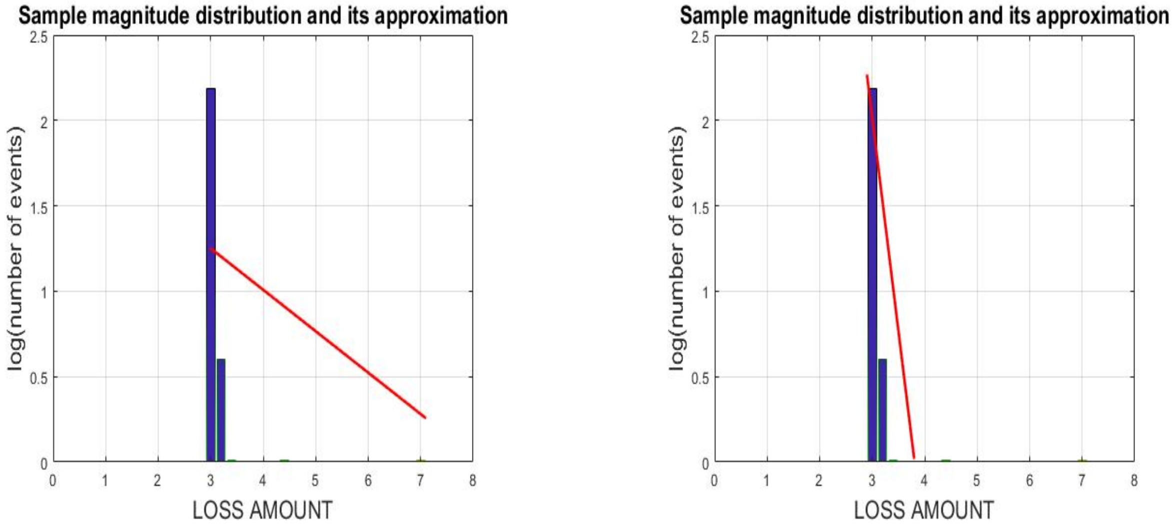
© University of Pretoria

Data split according to Loss Type



(a) *Method of moments* (b) *Maximum likelihood*

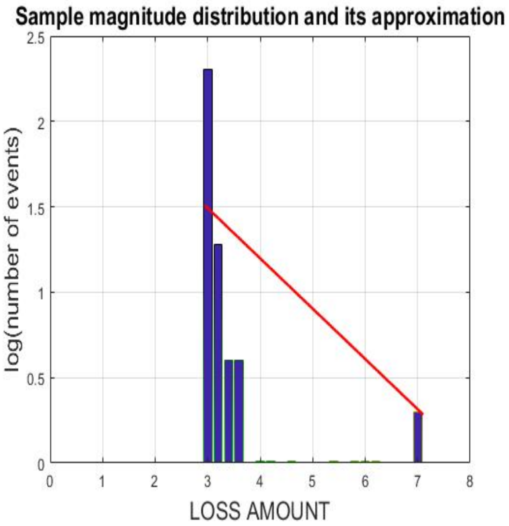
Figure 8.29: Loss type: Unauthorized Trading



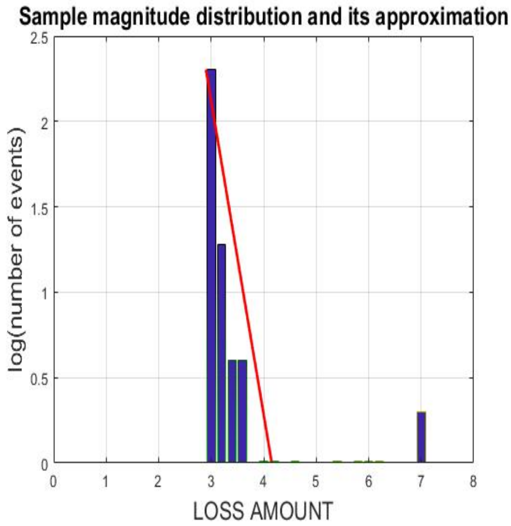
(a) *Method of moments* (b) *Maximum likelihood*

Figure 8.30: Loss type: Theft

© University of Pretoria

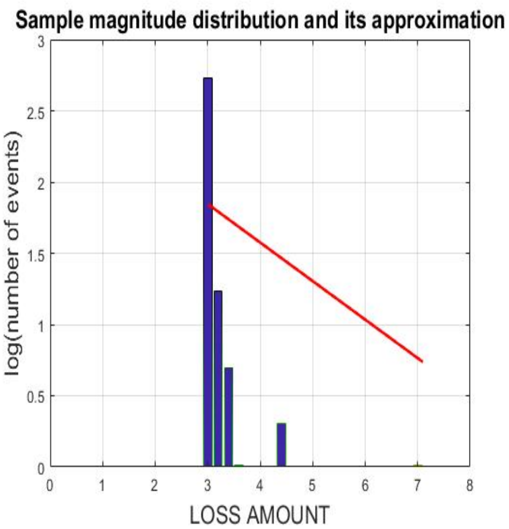


(a) *Method of moments*

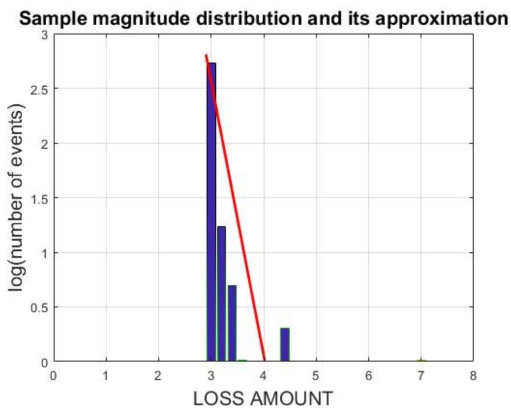


(b) *Maximum likelihood*

Figure 8.31: Loss type: Employee Relations



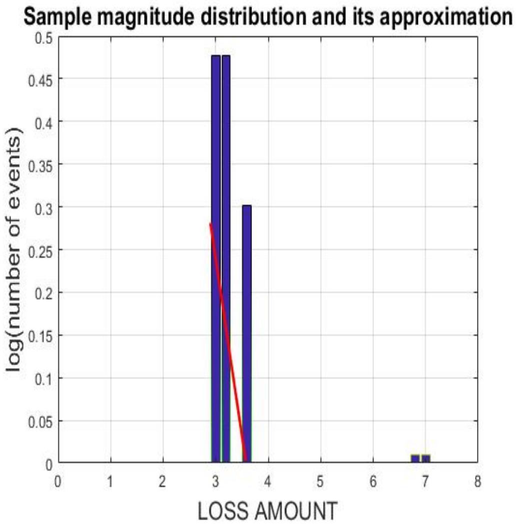
(a) *Method of moments*



(b) *Maximum likelihood*

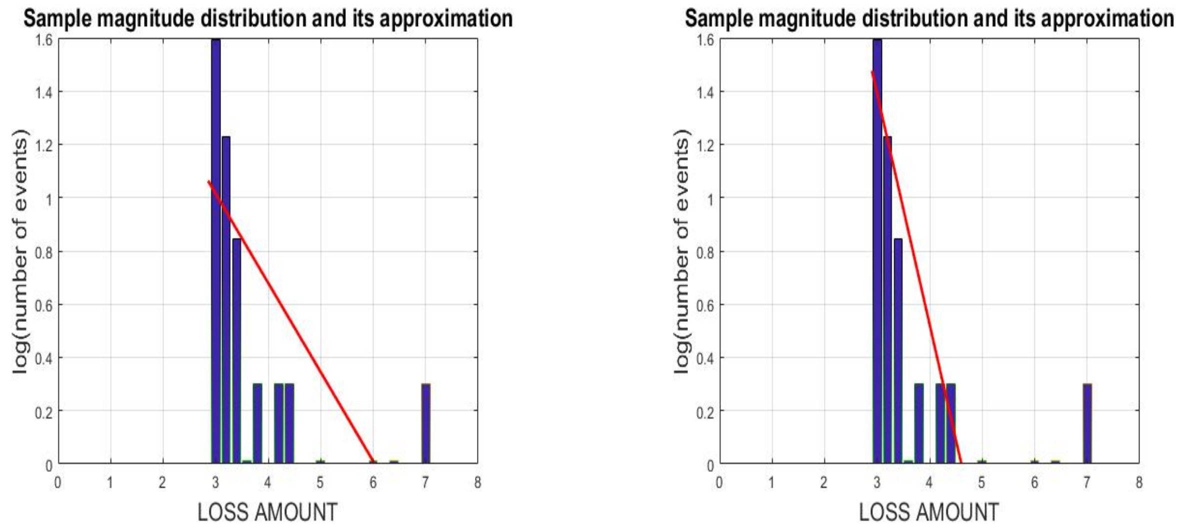
Figure 8.32: Loss type: Suitability, Disclosure Fiduciary

© University of Pretoria



(a) *Maximum likelihood*

Figure 8.33: Loss type: Business Disruption and System Failure



(a) *Method of moments* (b) *Maximum likelihood*

Figure 8.34: Loss type: Transaction capture, Execution and Maintenance

© University of Pretoria

Aggregate Losses

To generate aggregate losses, 100,000 Poisson loss frequencies where generated. Loss amounts where generated for each loss frequency for different α values and the resulting histograms were plotted and shown in Fig. 8.35 to 8.38.

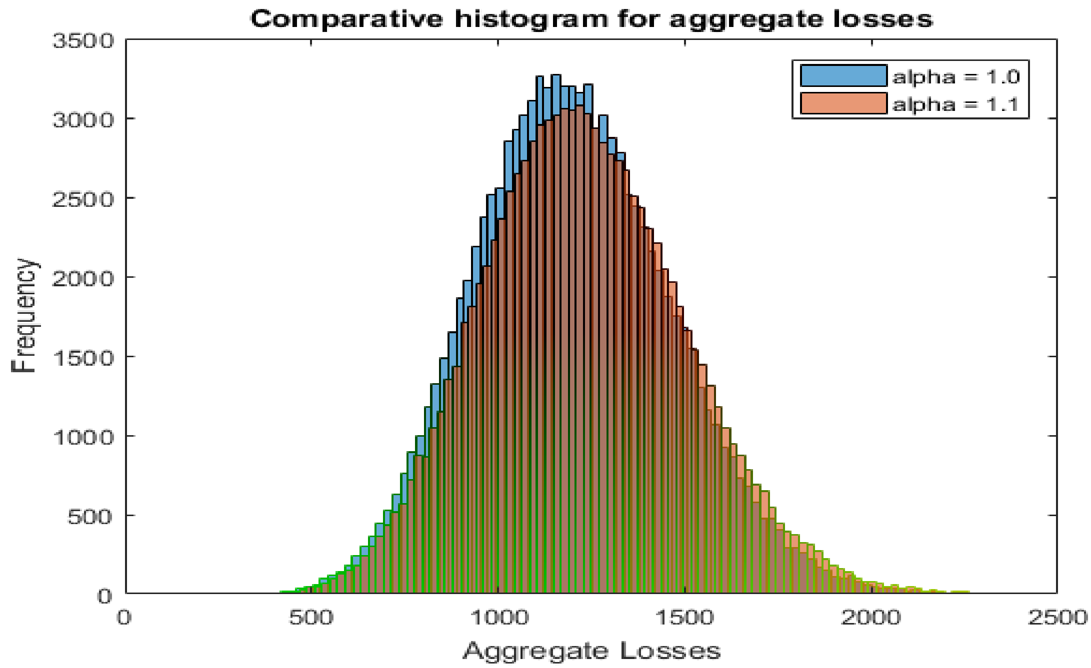


Figure 8.35: Aggregate losses distribution

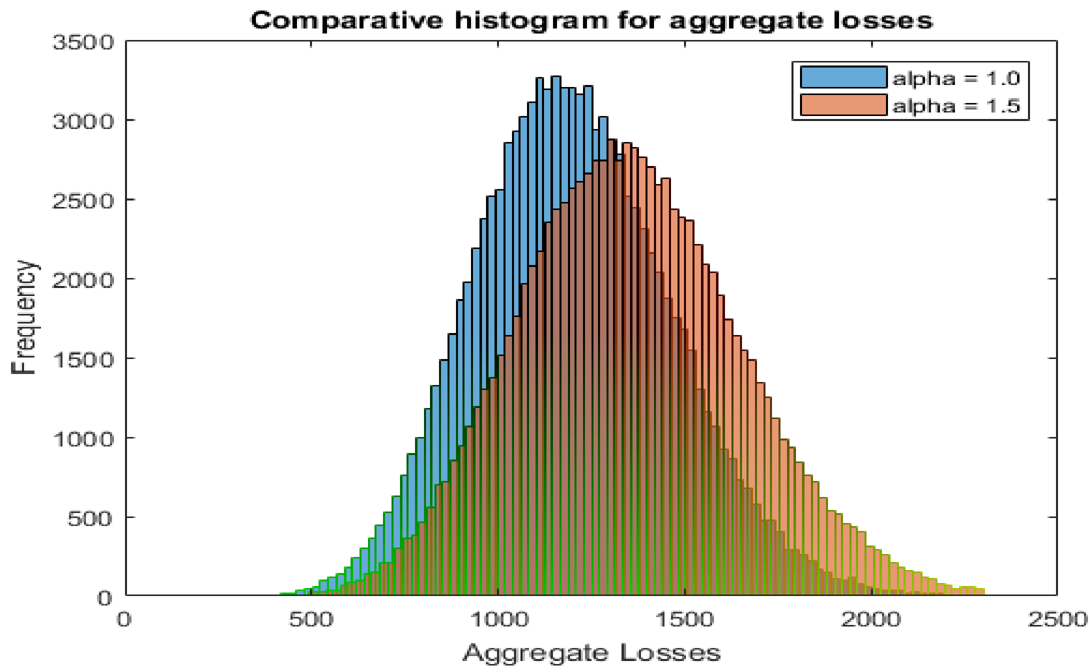


Figure 8.36: Aggregate losses distribution

© University of Pretoria

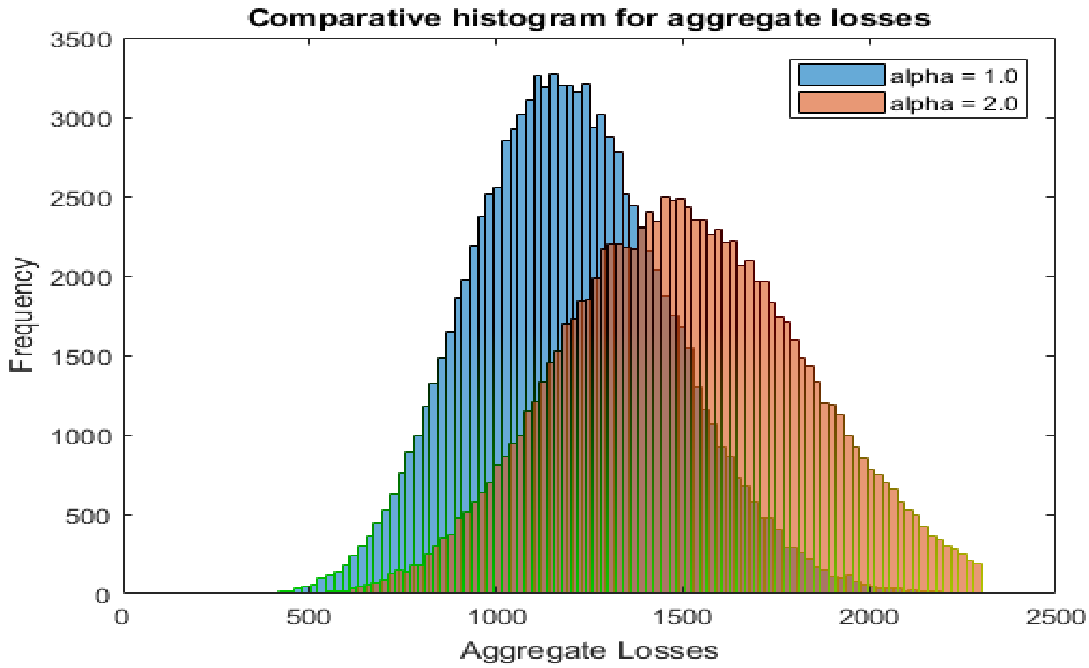


Figure 8.37: Aggregate losses distribution

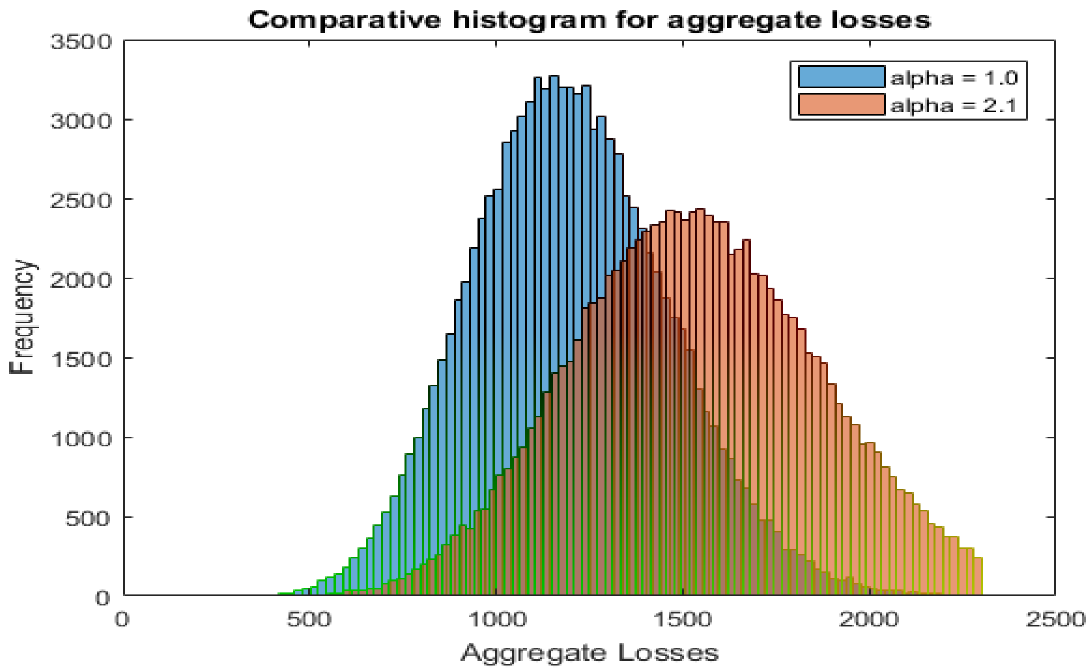


Figure 8.38: Aggregate losses distribution

© University of Pretoria