Table of Contents

## Figures



## Tables



## Equations



## Appendix

# 1. Introduction

*This chapter will give an introduction to the work behind this paper. This chapter will give background information to the problem this paper will address as well as talk about the purpose of this work.*

## 1.1 Background

Banks are today using quantitative models to calculate and analyze different types of risk. However, operational risk has proven to be difficult sometimes to use quantitative models on. The main issues are the uncertain nature of operational risk and especially the lack of historical operational loss data (Bakker, 2004). The result is that for some operational risk category's, it does not exist enough historical incidents that could be associated with this type of risk for a quantitative model to be used in a meaningful way.

Operational risk has historically been the *residual category for risk* (Power, 2003). Therefore, operational risk has become the "left-over" risk category for losses which cannot be related to financial risk or systematic risk. *Operational risk is treated as a left-over category from the cost banking risks* (Acharyya, 2012). The sub-categories that make up operational risk can, therefore, differ from each other in a more predominant way than for financial risk. This paper explores the possibility of constructing a more accurate quantitative model for estimating operational risk exposure by modelling each sub-category of operational risk individually. For simplicity reasons, this paper has only looked into one specific sub-category, namely IT risk.

The banking industry today is in a transition period where banks are quickly moving towards a digitalization of the banking processes (Broeders and Khanna, 2015). The result is that the operational processes within the banks are getting a lot more digital. The digitalization of banks has increased drastically in short period of time which has resulted in a heavy dependency IT systems have happened very quickly. Information technology and infrastructure is now a very important part of any financial institution. ECB wrote in a report that IT systems *continuity and resilience need to be sufficiently robust and tested to ensure timely recovery from operational disruptions*. And that *this is predominantly an area of concern for supervisors*. (ECB, 2016). However, bigger IT failures do not occur frequently enough to make a good

quantitative estimation of risk exposure. Smaller IT incidents occur a lot more frequent but have an insignificant impact on the banks bigger operations and could, therefore, be considered less relevant.

This paper will seek to estimate the risk of IT failure for the banking industry. This will provide an estimate of what this risk can amount to indirect cost for the whole industry. This can be a useful benchmark for the individual banks to put their own losses in relationship to.

### 1.2 Definition

**Operational Risk** is defined by the European Banking Authority (EBA) as *the risk of losses stemming from inadequate or failed internal processes, people, and system or from external events*. (EBA, 2017). It is clear that by most definitions, the IT risk is part of the operational risk and therefore regulated by the BASEL frameworks (Operational risk was first covered by BASEL II (BIS, 2016) and have been included in all later frameworks).

**VaR** or Value at Risk is a summary statistic of losses and the Financial Analysts Journal defines it as *a measure of losses resulting from "normal" market movements* and continues explaining that *Losses greater than VaR are suffered only with a specified small probability.* (Thomas J. Linsmeier and Niel D.). Usually, these losses are associated with 5 % or 1 % probabilities of occurring corresponding to $VaR_{95\%}$ and $VaR_{99\%}$ respectively. This paper will focus on $VaR_{95\%}$ which would, therefore, correspond to the worst losses possible with and 95 % accuracy. A higher level of VaR would make less sense given the specific risk being investigated, which is quickly changing. The banking industries exposure to IT risk will probably not look the same in 10 years which makes the highest VaR level perhaps a little too extreme.

**Monte Carlo Method** or a Monte Carlo Simulation defined by the book Monte Carlo Methods to be part of *the branch of experimental mathematics which is concerned with experiments on random numbers* (J.M. Hammersley and D.C. Handscomb, 1979). Hence, it is a process of generating and analysing random numbers.

**IT risk** or Information Technology risk is in this paper referred to the risk of having an IT system or infrastructure failure which interfere with a process within a financial institution. The Institute of Operational Risk defines the term "risk" as *something that has not yet caused a direct operational problem for the firm however there remains some degree of uncertainty concerning future outcomes* (The Institute of Operational Risk, 2010). This definition is taken to the more specific

IT risk which could be caused by a big variety of IT-incidents. Examples of IT incidents mentioned in this paper could range from website problems which prevent customers from accessing their accounts to problems with the internal cash-flow system which make customers unable to pay invoices in time, to as extreme as damage done by hacking- or cyber-attacks. All these incidents can cause costs for either the customers, which then often are compensated by the institution, or for the bank directly. This cost is referred to as direct cost and is the losses this paper are modelling.

### 1.3 Problem Statement

The aim of this thesis is to find an appropriate way of modelling IT risk for the banking industry.

### 1.4 Purpose

The goal of this thesis is to quantify the banking industries exposure to IT risk using LDA-based models and to evaluate which models are the best to use for this risk.

The reason for applying the quantification of this specific operational risk on an industry level is to find the banking industry's general exposure to the IT risk. It will be achieved using quantitative, LDA-based models which otherwise would be difficult to use on an individual banks level due to lack of statistical foundation.

The result of this paper will be a determination of which quantitative model that does the best job in estimating future yearly losses caused by bigger IT incidents. This quantitative model will use Monte Carlo method to simulate possible incidents and direct future cost which will form an aggregate loss distribution. The aggregate loss distribution will serve as the benchmark which the individual banks could use to evaluate their own IT risk exposure with.

### 1.5 Problem Description and Relevance

IT-systems are today generally very efficient and allow for example money to flow smoothly between institutions, people to check the balance on their accounts or withdraw or deposit cash. They play a major role in the functionality of our financial system nowadays. These IT-systems can, therefore, be very big and complex for larger financial institutions. All types of

IT-systems can experience technical difficulties from time-to-time. Such a failure within an IT-system can lead to some banking processes are being disrupted. Also, depending on what process is affected, it can lead to a quite high direct cost for the bank if this IT incident is not solved in time. These IT-systems are therefore exposing the banks to a risk which this paper refers to as IT risk. Indirect cost is also very predominant in this type of risk but extremely hard to quantify in a meaningful way. So for simplicity reasons, indirect cost is ignored in this paper and is left for the individual institutions to make their own estimations based on their individual goodwill and such.

This paper will try different quantitative LDA-based models and evaluate which one is the best to estimate the IT risk exposure for the banking industry. The problem this paper then approaches is the problem of insufficient data. Banks cannot usually make quantitative approaches, like LDA, to model IT risk since they usually lack enough internal data of this specific risk to make a good estimation. This paper solves this problem by including data from many different banks and other financial institutions in order to create achieve an estimation over the banking industry as a whole. This could be useful for the individual institutions to get a picture of what the aggregated losses look like and to compare their own risk exposure with it.

### 1.5.1 Distortion of Data: "Gap" in the Severity Distribution

Operational risk has been impeded by the lack of data (Fontnouvelle Rueff Jordan and Rosengren, 2003). When an incident occurs it can usually be solved very quickly by using alternative infrastructure or IT-systems to run the operational process while the faulty system or infrastructure is being handled. The result is that for the most part, these incidents are either without or with a very low direct cost for the banks. But if an incident occurs that cannot be solved in this way, it would mean a tremendous cost for the bank in question. The direct cost this incident could amount to is very much dependent on the process it interrupts and what problem it causes the bank. It can, therefore, vary from a very low amount to such a big amount it could jeopardize the bank's continued operations. Hence, it is a jump in the cost associated with each incident where most incidents end up without the cost and a few with a high cost. While these high-frequency, low-cost incidents can cause some minor headache and hidden indirect cost for some departments within the bank, it is usually of low importance from a risk management's point of view. These types of incidents are therefore ignored in this paper. The low-frequency, high-cost incidents that could happen are on the

other hand of higher importance and are the types of incidents which this paper will focus on.

### 1.5.2 Scarce Data

This led us to the next problem. Since these incidents rarely happen, it is hard to find enough recorded data for any meaningful analysis from an individual bank's point of view. This is indeed a problem for the IT risk but also a very common problem for operational risk in general (Jöhnemark, 2012). To tackle this usual problem of scarce data, the risk management team usually has two options; to choose a qualitative approach which relies more on expert inputs, or to make use of complimentary data to solve the lack of internal, in-house data (Bakker, 2004). Complimentary data can be scenario data, which is artificially generated data, or external data, which is data taken from other actors in the industry (Jöhnemark, 2012). The first mentioned qualitative approach is perhaps the most used in practice, although it might not be obvious. When dealing with an operational risk and the incidents are occurring too rarely or the data collection is not complete, the management will easily fall back on expert opinion or "gut-feeling" when evaluating the probabilities and severities of that specific risk. This could be just as accurate, or even more accurate sometimes, as a quantitative approach (The book: Foundations of Risk management (Aven, 2003) is recommended for deeper discussions regarding conditioned probabilities and working with heuristics). However, the problem with a qualitative approach is the lack of statistical evidence to back it up. There are essentially qualified guesses which could be biased (Ratner, 2002).

### 1.5.3 Dependence on External Actor's IT-system

It exists a high degree of inter-connectivity between banks in the financial industry today. For example, if an IT-system would interfere with the money flow of a certain bank, this could negatively affect other banks which were dependent on the money-flow of the previous bank. This cost caused by these incidents is not always recovered or compensated between institutions because of various reasons, including difficulties in proving the exact amount of the loss. The final result is that banks are not only very dependent on the functionality of their own infrastructure and IT-systems but also on the functionality of other banks infrastructure and IT-systems. This exposure to external IT-systems makes a quantification of IT risk on an industry level more relevant even for the individual banks.

# 2. Literature Review

*This chapter will discuss previous work done in this field. The chapter will discuss in detail the literature that this paper will be based on as "The Most Relevant Literature", as well as discuss more briefly other work that is relevant for this paper as "Literature Overview".*

## 2.1 Most Relevant Literature

In the article called Quantifying Operational Risk in Financial institutions, the Loss Distribution Approach is applied to quantify the operational risk of an anonyms US bank. The methodology, approach and difficulties are discussed. The article mentions that a drawback to this model is the difficulty in fitting a distribution to severity data and determining the distribution of the resulting aggregate loss (Keller and Bayraksan, 2011). This paper will adopt the same approach to calibrating the model and will therefore use the same methodology and approach as the article used. However, instead of looking at operational risk as a whole, this paper will focus more specifically on a single source of operational risk, namely the IT risk. The idea is that the problems mentioned by Brian Keller and Güzin Beyraksan in this article could be overcome by focusing on a more specific IT risk.

In a master thesis called Modelling Operational Risk, where the operational risk were discussed regarding how to model it. The goal is to try different ways of modelling operational risk and find the best-fitted one to use for financial institutions who are using the advanced measuring approach. The research concluded that a Compound Poisson distribution is best suited for modelling frequency and that severity distribution is best modeled by a piecewise defined distribution with an empirical body and generalized Pareto tail. This paper will only focus on one specific source of risk under operational risk, namely IT risk. However, this paper will use the same methodology as Jöhnemarks work and see if the same distributions hold true for even for this sub-risk. Alexander Jöhnemarks master thesis is an important work for which the work in this paper is based on. The different methods of modelling operational risk which has been tried in Jöhnemarks thesis are of special interest and this paper will use these methods when modelling IT-risk.

In a report called Using Loss Data to Quantify Operational Risk, it is suggested that operational risk is an important risk for banks and the capital charge will often exceed the

charge for market risk (Fontnouvelle Rueff Jordan and Rosengren, 2003). Just like the other literature sources this also seeks to find an appropriate model for quantifying operational risk. However, what makes this report interesting is that it includes external, publicly available data. They also discuss the possible problems with bias data, referring to a positive correlation between the likelihood of an incident being reported and the amount of severity the incident inflicts. The data-sampling problem is very likely to exist in many types of operational risk and is something to consider from an individual banks perspective. This paper will continue the discussions of the use of external data in operational risk modelling but will focus on one specific risk, the IT risk. When it comes to the problem of biased sample, which would contain a *disproportionate number of large losses* (Fontnouvelle Rueff Jordan and Rosengren, 2003), this problem is not considered to be as significant in the data this paper is using. Since this paper is only focusing on incidents of significant value. Furthermore, since banks are required to report losses from operational risk the data can be assumed to be a random sample accurately representing the population of IT incidents of significant value. That being said, the incidents which amount to lower values in this data can still be subject to this problem mentioned in Fontnouvelles, Rueff, Jordan and Rosengren's work.

A report, published by the bank UniCredit, called R and Operational Risk shows how to use AMA models in R (Piacenza, 2012). This report contains instructions for how to mathematically construct an AMA model and run Monte Carlo simulations in R as well as displaying detailed examples of such models and their output. This work and instructions have been carefully considered and influenced the mathematics behind method two in this paper. The main usage of this report has been for the construction of the hybrid severity distribution. Please note that while Piacenza's report used R as an analytical software, this paper have used @Risk from Palisade.

### 2.2 Other Relevant Literature

Besides the most important work for this paper which has been presented above, there is plenty more academic work published about modelling techniques for operational risk which cover interesting theoretical approaches that are still relevant for this paper. Most models cover operational risk as a whole and never focus on creating a model for a single underlying source of risk, like IT risk. However, there is good variety of work which includes operational

risk with many data points and others with low data points. In this paper, the following work should be highlighted:

- Fundamentals of Risk Analysis: A knowledge And Decision-Oriented Perspective is a book which discusses thoroughly how to approach model-building, how to think about uncertainties and how to use risk analysis in decision-making processes. The book has provided a framework for conducting and understanding risk analysis, suitable for finance as well as other fields (Aven, 2003).
- Quantifying Operational Risk within Banks According to Basel II is a master thesis which introduces a method for quantifying operational risk which complies with the Advanced Measurement Approach (AMA). How to work with risk modelling is discussed with a specific focus on risk with the low amount of data. This specific paper solves the problem of low quantitative data with a so-called LEVER method where the internal data are complemented with artificial qualitative data. This paper makes use of the same idea and method of complementing the lack of data. However, the use of the LEVER method is not considered to be as well applicable to IT risk (Bakker, 2004).
- LDA at Work is a published paper presenting the capital model for Deutsche bank. Deutsche bank follows the Loss Distribution Approach which is a common approach within the AMA. This work shows how to make use of loss data in severity and frequency modelling and also discusses the implementation of dependence. It also explains the capital calculations used in LDA. This is a very relevant work for this paper and this method of finding the right model in this work has influenced the approach in this paper (Aue and Kalkbrener, 2007).
- The Quantitative Modelling of Operational Risk between G-and-H and EVY brought up a thorough discussion about the newly proposed parametric g-and-h distribution by Dutta and Perry, which were supposed to act as an alternative model for quantification of operational risks with the lower dataset. The work also discusses the link between the g-and-h distribution and the extreme value theory. The conclusion of the work showed that the quantile estimation, using extreme value theory, could lead to inaccurate result when the data are modeled by g-and-h distribution (Degen Embrechts and Lambrigger, 2007).

- A Bayesian Approach to Modelling Operational Risk When Data is Scarce (Svensson, 2015) is a thesis which tried to create an AMA model for operational risk where internal data is very low. Just like this paper will, K. Petter Svensson tried to solve the lack of internal data by including external data (as well as scenario data). Different from other work, this thesis concluded that it is "possible to build an AMA model with Poisson loss frequencies using Bayesian inference to combine the different data sources" (Svensson, 2015). Svensson's dissertation used AIC and BIC score to find the most suitable distribution in their model which is a technique this paper will use as well.

# 3. Theory

*The purpose of this chapter is to explain the underlying theory behind the methodology in this paper.*

## 3.1 Qualitative or Quantitative Approach

Any research approach can be generalized to follow either a qualitative or quantitative approach. The qualitative approach has its advantages as it is possible to get a more in-depth analysis (Gill Stewart Treasure and Chadwick, 2008) which will often generate soft-value results. This paper will apply a quantitative approach which focuses on a broader number of participants and often applies statistical techniques. A drawback with the first mentioned qualitative approach to risk analysis is the fact that it requires a lot "guesswork" which makes the estimates less reliable (Bakker, 2004). This is the upside of the quantitative approach instead. However, the quantitative approach requires a larger amount of statistical data instead. Data that is not so common with certain types of operational risk, like the IT risk.

## 3.2 Regulatory Framework

When dealing with risk one typically deals with estimated cost and probabilities. All probabilities are conditioned on the background information (and knowledge) that we have at the time we quantify our uncertainty (Aven, 2003). Many operational losses happen frequently and do not result in major damages. These include everything from small data entry mistakes to minor system failure. However, banks (as well as other financial institutions) can suffer from the operational risk that can cause major losses which are of great concern for a risk manager. It is, therefore, paramount for banks to protect themselves from losses due to operation risk than show the range and magnitude of this risk (Keller and Bayraksan, 2011), and this includes IT risk. Since Basel II was finalized in June 2006 the banks were required to calculate the capital need to cover losses due to operational risk. The Basel II accord allows three ways of calculating operational risk. These are the Basic Indicator Approach, Standardized Approach, and Advanced Measurement Approach. The Advanced Measurement Approach, fourth ward denoted as AMA, allows the banks to develop their own model for estimating their operational risk exposure. The AMA models are usually more complex than the basic indicator or standardized approaches. However, the AMA model usually typically yields better estimates of risk (Keller and Bayraksan, 2011). The bank must have its own, in-house developed model approved first by the respective authority. Dr. Pavel V. Shevchenko's book; Modelling Operational Risk Using Bayesian Inference. The Loss

Distribution Approach is one of the most commonly used models under the AMA according to multiple studies including Keller and Bayraksan (2011), Franchot Georges and Roncalli (2001) and Shevchenko (2011). The Loss Distribution Approach, or LDA, is the model this paper will use when quantifying the IT risk for the banking industry.

The Bank for International Settlements (BIS) is actively working to withdraw the opportunity to exercise the advanced measurement approach for calculating the bank's capital requirement for operational risk (BIS, 2016). However, this paper will use LDA to calculate an industry's exposure to a certain type of operational risk, not to give a specific actor in this industry any suggestion on the capital requirement. Therefore, using the LDA method is still interesting and would generate a good estimate of the risk.

### 3.3 Loss Distribution Approach Model

The LDA model needs statistical data of a risk in form of yearly frequency, of which an event occurs, and the monetary value of the losses (severity) given that an event occurs. These two are assumed to be independent of each other and modeled separately (Svensson, 2015). A relevant distribution is fitted into the yearly frequency and the loss, which in turns are being used as inputs to calculate the aggregate loss distribution. To obtain the aggregate loss distribution it is common to use a Monte Carlo simulation. In this paper, the LDA method will be used accordingly. This means that the yearly frequency and losses given an incident will be measured and fitted to an appropriate distribution. A Monte Carlo simulation will then be used with these distributions as input to generate an aggregate distribution for this risk on an industry level.

The LDA is used when modelling the IT risk in this paper. This approach was chosen because it is a quantitative approach which otherwise would have been hard to use for an individual bank to estimate this risk (because of the previously mentioned problem of scarce data, see section 1.5.2: Scarce Data). The reason to why this paper used LDA as a quantitative method and not any other quantitative method is because it is one of the most popular methods under AMA (Shevchenko, 2011). AMA allows the bank to build its own, in-house model for quantifying its operational risk exposure. And since LDA is one of the most used methods in the industry for banks who create their own models, it is probably the best-suited model for quantifying this IT risk.

## 3.4 Simulation Method and Distribution

This paper is going to use historical incident data from the banking industry to find and fit appropriate frequency and severity distributions. These distributions will be used as input in a Monte Carlo simulation in order to estimate the aggregate loss distribution of the IT risk. The frequency distribution will be corresponding to the number of incidents that occurs in a given year and will, therefore, be following a discrete distribution. The severity distribution will be corresponding to a number of losses experienced by the industry given an incident occurs and will, therefore, be following a continuous distribution. The distributions will be fitted from historical data using risk analysis software @Risk from Palisade. How well the distributions fit the data will be determined by Akaike information criterion (or AIC), and Bayesian information criterion (or BIC).

### 3.4.1 Akaike Information Criterion (AIC)

AIC is a measurement of relative quality of a statistical distribution for a given set of data and is something the risk analysis software will help determine. The AIC measurement is based on information theory and will indicate how much information is lost from the data if the given distribution is assumed, in relationship to the other models. The best model is, therefore, the one which minimizes the AIC score (Liddle, 2008). AIC is calculated according to the following formula:

*Equation 1 - Akaike Information Criterion*

$$AIC = -2\,Ln\,\mathcal{L}(\max) + 2k$$

where $\mathcal{L}$ (max) is the maximum likelihood achievable by the model and k is the number of parameters in the model (Liddle, 2008).

### 3.4.2 Bayesian Information Criterion (BIC)

The Bayesian information criterion, or BIC, was introduced by Schwarz and it assumes that the data points are independent and identically distributed (Liddle, 2008). BIC works in the same way as AIC, namely, it will rank the best-fitted distribution according to a BIC score where the lowest value will be the best-fitted distribution. According to the website: standfordphd.com, BIC has a preference for simpler models, with a lower number of parameters, than compared to AIC (Standfordphd.com, u.d.). BIC is calculated according to the following formula:

*Equation 2 - Bayesian Information Criterion*

$$BIC = -2\ Ln\ \mathcal{L}(\max) + k\ Ln\ N$$

where N is the number of data points used in the fit (Liddle, 2008).

### 3.4.3 Monte Carlo Simulation

The distributions found to be a good fit for the historical losses will later be used in a model. One distribution is used for modelling frequency while one or two distributions are used for modelling severity. These two or three distributions are used as inputs in Monte Carlo simulations. A Monte Carlo simulation is an open form solution which could be done in multiple ways but involves solving analytical formulas by using a large quantity of randomly generated numbers. (Navarrete, 2006).

## 3.5 Probability Distribution

Probability distributions are defined by a probability function which assigns the probabilities to the possible values of the random variable (Jones, 2017). Hence, a probability distribution lists the possible outcomes of a random variable together with its corresponding probability.

In most general terms, a probability distribution can be seen as a discrete probability distribution or as a continuous probability distribution. It is the values that the random variable can assume that determine this and is a central subject of the probability theory (Andale, 2017). If a random variable can only assume a finite number of values, it would be a discrete distribution and if the random variable could assume an infinite number of values, it would be a continuous distribution. However, there are more ways the many different distributions are categorized and one common way is by looking at their parameters. Many distributions are not a singular distribution but a family of distribution. (Handbook Engineering Stastistics, 2017). It can depend on if a distribution have one or more shape parameters. The shape parameter allows a distribution to take on a variety of shapes, depending on the value of this parameter (Handbook Engineering Stastistics, 2017). A family of distributions includes distributions who are sharing some properties or characteristics. When describing the distributions used in this paper, some common family of distributions are used. The exponential distribution family is one of the most common distribution and

includes many of the commonly used distributions. Many of the distributions used in this paper belong to this family. Clark and Thayer (2004) introduces the exponential family in their paper explaining how they are suitable for aggregate loss models. However, other some distributions are included that do not belong to this distribution and belongs to other distribution families instead. An example of a less common family of distributions would be the Pearson family, which are *characterized by two quantities usually referred to as $\beta_1$ and $\beta_2$* (Lahcene, 2013).

This paper sought to model IT risk using certain distributions to explain the data. This was done by an analytical software where many different distributions where included. However, only a few number of distributions were suggested and later implemented in the models. The theoretical background of the distributions who were included in the models of this paper are explained later in this chapter under the subheadings: "Discrete Probability Distribution" and "Continuous Probability Distribution".

## 3.6 Discrete Probability Distribution

Discrete distributions are used to model frequency. This paper uses discrete distributions to model the number of incidents which occurs within a year. The result of these distributions will hence be a distribution of all possible incidents that could occur in an upcoming year.

### 3.6.1 Poisson Distribution

The discrete Poisson distribution is a probability distribution of the random variable X. This distribution describes the probability of a certain number of events occurring, usually expressed as k, within a given range (Frost, 2017). The Poisson distribution is a member of the exponential family and includes a parameter describing the expected number of events occurring denoted as lambda (Clark and Thayer, 2004). The probability density formula for this distribution is the following:

*Equation 3 - Poisson Distribution's Probability Function*

$$f(k|\lambda) = \frac{\lambda^k}{\mathrm{k!}}\ \mathrm{e}^{-\lambda}$$

### 3.6.2 Negative Binominal Distribution

The Negative Binominal distribution belongs to the Exponential distribution family and is a discrete probability distribution based on two parameters (Clark and Thayer, 2004). It is a distribution of the number of successes in the sequence of Bernoulli trials before a specified number (r) of failures, and the success probability (p). The probability density formula looks like this:

*Equation 4 - Negative Binominal Distribution's Probability Function*

$$f(k) = \binom{r+k-1}{k} p^r q^k$$

## 3.7 Continuous Probability Distribution

This paper is using non-negative continuous distributions to model the severity. The distributions used in this paper have been calibrated so they cannot assume any negative numbers. This has been done since operational risks, like the IT risk, can only assume a loss for the company if they occur, unlike financial risk for example. Continuous distributions are used to model a number of losses caused by a bank given that an incident occurs. Unlike the frequency modelling, the severity of an incident can amount to a non-integer. Hence the random variable, which is the severity in this paper, can take an infinite set of values.

### 3.7.1 Pearson 5 Distribution

The Pearson 5 Probability Distribution is a three parameters, continuous probability distribution belonging to the Pearson distribution family (Lahcene, 2013). The Pearson typed distributions are characterized by two quantities commonly referred to as $\beta_1$ and $\beta_2$ (Lahcene, 2013). The probability density formula for this distribution is as following:

*Equation 5 - Pearson 5 Distribution's Probability Function*

$$f(x) = \frac{\exp\left(-\frac{\beta}{x-\gamma}\right)}{\beta\Gamma(a)\left(\frac{x-\gamma}{\beta}\right)\alpha^{+1}}$$

### 3.7.2 Log-Logistic Distribution

The Log-Logistic Probability Distribution is a continuous distribution of a variable whose logarithm has the logistical distribution. This distribution belongs to the Logistic distribution family for example (R-forge distribution Core Team, 2009). The log-logistic distribution can, in practice, be used as an alternative to the lognormal distribution (Hamedani, 2000) which shows the similarities of these two distributions. The probability density formula for the Log-logistic distribution is the following:

*Equation 6 - Log-Logistic Distribution's Probability Function*

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

### 3.7.3 Inverse Gaussian Distribution

The Inverse Gaussian Probability Distribution (also known as Wald or normal-inverse Gaussian distribution) is a two parameters continuous distribution (Andale, 2017) which also belongs to the exponential family (Clark and Thayer, 2004). The probability density formula for this distribution is the following:

*Equation 7 - Inverse Gaussian Distribution's Probability Function*

$$f(x) = \left(\frac{\lambda}{2\pi x^3}\right)^{\frac{1}{2}} \exp\left(\frac{-\lambda(x-\mu)^2}{2\mu^2 x}\right)$$

### 3.7.4 Lognormal Distribution

The Lognormal Probability Distribution is a continuous probability distribution of the random variable X, whose logarithm is naturally distributed. The result is a distribution which is skewed to the left. This distribution is a member of the general exponential family (Clark and Thayer, 2004). The probability density formula for this distribution is the following:

*Equation 8 - Lognormal Distribution's Probability Function*

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

These discrete and continuous distributions are the distributions used in this paper to model IT risk. Even though not every distribution belongs to the same family of distribution, all continuous distribution are similar in the way the show a positive skew. The analytical software included more distributions in the search for the most suitable distributions to model the given data. The distributions included in this testing process can be viewed in the method chapter 5.1: Fitting the Model. The distribution deemed to be the best fit are the distribution mentioned above. The distributions considered to not be a good enough fit by the software were not used later in this paper and therefore did not have their theoretical background explained in this chapter.

# 4. Data

*The purpose of this chapter is to inform the reader about vital information regarding the data used in this paper. Here will be discussed the qualities and limitations of the data.*

## 4.1 External Data

To tackle the problem of low data this paper is going to use external data for fitting the frequency and severity distributions to. Meaning the result will be a quantification of IT risk in the European banking industry. External data could be used, from an individual bank's perspective, in situations where the source of risk is not very unique to the single organization. However, even though the data is taken from the same population there could be different value-criteria (or threshold) for when incident are reported by different institution (Mignola, 2003). This problem is approached by incising the threshold over which data to include in the distribution-fitting. This paper only focuses on IT risk of significant direct costs which in the problem of different thresholds are minimized.

The external data corresponds to the same categories of incident which has occurred by other financial institutions in the industry, sometimes also referred to as public data (Guillen Gustafsson Nielsen and Pritchard, 2007) (even though the data this paper uses is not public). When an individual actor uses external data, it is important to make sure that the conditions for the risk are relatively similar to the industry which the complementary data are taken from. For example, it would not make much intuitive sense to use external data to estimate the risk of fire occurring in an office. Although the individual bank may have a few fires occurring from time-to-time, other financial institutions might be completely digitalized and cannot, therefore, have the same problem with fire incidents. So the industry risk of a fire occurring will in this case not represent the individual bank's own risk of fire occurring. However, all banks and financial institutions are today digitalized to some extent and therefore rely on IT-systems to work properly. All banks in the banking industry have both internal and external banking processes which heavily rely on the functionality of IT-systems and IT infrastructure. It can, therefore, be concluded that this industry risk will be relevant for the individual banks within this segment. In fact, not including external data in some circumstances, like this one, could lead to an underestimation of the severity of rare events. Internal data should be supplemented with external data *in order to give a non-zero likelihood to*

*rare events* which could be the case if only internal data are considered (Frachot and Roncalli, 2002)

This type of risk is more homogenous between the banks. No bank has yet managed to find any completely flawless IT systems which are without any incidents. However, the policy for managing the risk and maintaining the IT systems, as well as the level of skill with the people working with these systems might not be homogenous. It is, therefore, important to keep in mind that the result of this paper will be a quantification of the industry's IT risk, which might not be a perfect representative for the individual bank's IT risk. It will, however, work as a benchmark to the industry risk which an individual bank could use to put their own risk exposure in relationship to. This will give some indication for the risk management about the performance of their IT-systems when they can benchmark it to the industry total. A well working IT and infrastructure systems are getting more and more important in today's world.

This paper will make use of Monte Carlo simulations to generate the resulting aggregate loss distribution. Hence, the result will be scenario data (or output) which is characterized by being forward-looking (Jöhnemark, 2012).

### 4.2 Characteristics of the Data

This paper is using data from incident reports from all IT-system incidents which the banking industry experienced. The data is not published in this paper due to confidentiality. Because the data is confidential, it has therefore been anonymized and multiplied by a secret factor before it was used in order to keep the confidentiality of the data. This is very important to consider when viewing the resulting aggregated yearly loss. Because the data has been multiplied in this way, it should no longer be viewed as a monetary value but simply a number. What is published in this paper is only the simulated output of the aggregated yearly loss which is generated from distribution assumptions which were made from observing the actually multiplied data. The result published in this paper is therefore not to be viewed as the industry's IT risk exposure in monetary terms. However, the aggregated loss distributions can still be used to view the banking industry's risk exposure in relative terms. The focus of the paper will be on how to properly model this type of risk using the LDA method. The data can be assumed to be a random sample of the whole European banking industry's IT incidents.

The Data used in this paper dates back 10 years, from 2007-01-01 to 2016-12-31. Basel II was the framework which required the banks to start reporting and quantifying the operational risk. Since this framework was published in June 2004 (Bank for International Settlements, 2004), the financial institutions can be assumed to have started recording the operational risk of IT-system failure by 2007. One can argue that a proper reporting of operational risk did not get implemented immediately because it might have taken time to find functional processes. However, since it was required to report operational incidents of the relevant amount by 2007 it can be assumed that the data this paper is using are correctly representing the IT incidents financial institutions are experiencing. It is impossible to check if certain incidents have systematically been left out and resulted in a non-random sample in the database. For simplicity reasons, assumption has to be made that the data in this database is representing a random sample.

The data in this external database, after it has been multiplied by a factor, is referred to as the external data and correspond to incidents reported from the banking industry in the European region. This data is inspected and any abnormalities are analyzed and disregarded if it can be considered to be associated with a reporting error. All obvious duplicates of incident reports have been removed and the remaining cost of each incident have been adjusted for inflation (where the historical cost are adjusted to 2016 price levels for better comparison).

# 5. Method

*This chapter will discuss about the methodological approach which was used in this paper. This chapter will also explain how this quantitative models have been constructed.*

## 5.1 Fitting the Model

When building the model to estimate the future yearly losses from IT risk we first need to fit a frequency distribution and a severity distribution which can be used in a simulation of future events. To find the best distributions for this model the risk analytical Software, @Risk from Palisade, is used on the external data to find the best fitted distributions for both the frequency and severity. The software will find and calibrate, the most suited distribution based on the AIC score and BIC score. The most common distributions are tested for. The following discrete distributions are included when searching for the best fitted frequency distribution and the two most suitable distributions are later used in LDA models and are also explained in Section 3.5: Discrete Probability Distributions.

- Binomial Distribution
- Geometric Distribution
- Hypergeometric Distribution
- Uniform Distribution
- Negative Binomial Distribution
- Poisson Distribution

The following continuous distributions are included when searching for the best fitted severity distribution and the four most suitable distributions are later used in LDA models and are also explained in Section 3.6: Continuous Probability Distributions.

- Beta Distribution
- Chi-square Distribution
- Exponential Distribution
- Extreme value Distribution
- Gamma Distribution
- Inverse Gaussian Distribution
- Laplace Distribution
- Levy Distribution

- Logistic Distribution
- Lognormal Distribution
- Normal Distribution
- Pareto Distribution
- Pareto 2 Distribution
- Pearson 5 Distribution
- Pearson 6 Distribution
- Student's t-distribution
- Triangular Distribution
- Uniform Distribution
- Weibull Distribution

This paper uses multiple modelling techniques in order to find the most optimal model for this specific risk. All severities are assumed to be independent of each other and identically distributed.

### 5.1.1 Method One

The first method that was tried were also the simplest one. The frequency (number of incident per year) and severity (cost of a given incident) were both modeled by a single frequency and severity distribution respectively. The most relevant distributions were later used as input in a Monte Carlo simulation to estimate the aggregate loss.

Since multiple distributions had a similar fit to both the frequency and severity of the data, more than one distribution were tried as input for modelling frequency and severity. Each and one of the frequency distribution was tried with every severity distribution in a Monte Carlo simulation where the resulting aggregated yearly loss is then analyzed regarding their accuracy to the external data.

Here is a description of how the model works. First, the frequency distribution simulates a discrete number of times an IT incident would occur in a future year. Then, given the number of incidents which were simulated to occur, the model generated a simulated direct cost for each and one of those expected incidents which are based on the continuous severity distribution. The different cost of each incident was later summed up to become the

aggregated total loss of IT incidents during a year. This aggregated total loss is the output of the model which were repeated 10 000 times in a Monte Carlo simulation in order to create to aggregate loss distribution.

### 5.1.2 Method Two

Since the severity distributions showed a rather poor fit a new method was tried where the use of a mixed model was tested. The severity was modeled by a different body and tail distribution, both of continuous nature. Since the data contained a lot of extreme outliers in the right tail, the idea was to find a better-fitted model if the body and right-tail were modeled by separated distributions. When constructing the model used in method two, the instructions given in a report produced by UniCredit called R and Operational Risk, were closely followed. Specifically for merging the tail and body distribution in the severity modelling (Piacenza, 2012). Finding the threshold between the body and tail in the external data was done by visual identification. Since the external data did show a clear transition from body to tail values, this method was chosen because of simplicity. The body corresponds to approximately 96,05 % of the severity data while the tail where modeled by the remaining 3,95 % of the external values allocated to the right tail. The selection process for how the tail or the body distribution are chosen for a given incident is very much influenced by the method used in Fabio Piacenza in his work.

Here is a description of how this model works. First, the frequency distribution generates first a discrete number of incidents that would occur in the next year. Then, given the number of incidents which are simulated to occur, the model generates two simulated direct costs for each and one of those expected incidents which are based on the continuous body and tail distribution respectively. Both the body and the tail distributions are modeled separately by a best fittest distribution for the body and tail data separately. Next to these costs are other random number generated which will take any number from 0 to 1, where every number within this range has an equal chance of occurring. This number is then compared to a new parameter called Fu, which is essentially the threshold between tail and body distribution, namely 0,9605. If the random number would be below the Fu parameter, then the loss simulated from the body distribution is assumed. If the random number would be equal or bigger than the Fu parameter, then the loss simulated from the tail distribution is assumed. Each individual incident is generated this way and later summed up to become the aggregated total loss of IT incidents during a year. The aggregated total loss is the output of the model

and the process of obtaining this aggregated total loss is repeated 10 000 times in a Monte Carlo simulation in order to create to aggregate loss distribution showing all possible outcomes and their corresponding probability.

The threshold or barrier for this double-distribution model corresponds to around 96,05 % of the data.

The same frequency distributions used in Method one were used in Method two as well. The severity distributions used in method one were also used as body distributions in method two but recalibrated to fit only the body data. The best-fitted tail distributions were found using the same analytical software @Risk from Palisade where two distributions showed good fit. The two tail and four body distributions were combined in all possible ways which generated eight different "Hybrid" severity distributions. These eight hybrids were combined with the two frequency distributions which resulted in 16 different models.

### 5.2 Aggregated Loss Distribution

The aggregated loss distribution is estimated using the result from the Monte Carlo simulation. The estimation of the Aggregate loss for a given year will be calculated using the frequency distribution to estimate the number of an incident occurring in a future year and the severity distribution to estimate what direct cost an incident would impose. In Method two the severity distribution will be made up of a hybrid distribution. The aggregate loss would be the sum of a yearly simulated cost caused by simulated number of IT incidents. The Aggregate loss distribution would be obtained by repeating these aggregate loss calculation a large amount of times in a Monte Carlo simulation. This paper use 10 000 trails in the Monte Carlo simulations in order to get a large enough sample to estimate an aggregate loss distribution.

This paper will run a Monte Carlo simulation for all severity and frequency models proposed from each method. But the most accurate aggregate loss distribution will be discussed and analyzed based on how well the model fitted the data and how feasible the result turn out to be in comparison to the historical data and trends in the industry.

# 6. Result

*The purpose of this chapter is to present how the models where created in terms of what distributions they consisted of and how well the fit was. The simulated result of these models are also presented here.*

## 6.1 Method One

### 6.1.1 Frequency Distribution

The frequency distribution was obtained by fitting a discrete distribution to the historical data, which represent the number of incidents which occur within a year. The risk analysis software suggested two distributions which were all relatively close to the actual data distribution. The Negative Binominal distribution showed to be the best fit for the frequency with an Akaike information criterion (AIC) of 125.14 and Bayesian information criterion (BIC) of 124.03. The second best fit was the Uniform distribution with an AIC score of 125.76 and BIC score of 124.65. Finally, a third best fitted distribution was the Poisson distribution with an AIC score of 323,58 and BIC score of 323,38. However, since the yearly losses seems to be occurring with a preference to the median, the uniform distribution could be misleading. Only the Poisson and Negative Binominal distributions are considered since these both provide a higher probability for the frequency generated to be around the median in the tails.

*Table 1 - The Frequency Distributions AIC and BIC Scores*

| Discrete Distribution: | AIC | BIC |
|---|---|---|
| **Poisson** | 323,58 | 323,38 |
| **Negative Binominal** | 124,64 | 124,65 |

### 6.1.2 The Severity Distribution

The continuous severity distribution is obtained by fitting a discrete distribution to the historical data. These distributions showed a lot higher AIC and BIC scores, meaning that they do not fit as good as the frequency discrete distributions. However, this has to do with the number of data points included when calculating the AIC and BIC scores as well as the larger number of distributions used in the comparison. The discrete frequency distributions only included 10 data points (number of yearly incidents during a period of 10 years). The continuous severity distributions, on the other hand, included a lot more data points which

naturally inflated the AIC and BIC scores. The continuous severity distribution contains the cost of each and one of the incidents which occurred during this 10 year period which is so much more data points than the frequency distribution so the AIC and BIC scores are incomparable. The frequency and severity distributions must instead be looked at individually and can only be compared with a distribution of the same category.

The best-fitted distribution suggested by the software was the Pearson 5 distribution with an AIC score of 108 984. The top fitted distribution and their AIC and BIC score can be seen in the table below

*Table 2 - The Severity Distributions AIC and BIC Scores in Method One*

| Continuous Distribution: | AIC | BIC |
|---|---|---|
| **Pearson 5** | 108 984 | 109 002 |
| **Log-Logistic** | 109 466 | 109 484 |
| **Inverse Gaussian** | 109 870 | 109 888 |
| **Lognormal** | 109 883 | 109 901 |

Distributions that scored beyond this point had an AIC and BIC score which were very large in relation to first four and were therefore ignored. These four continuous distributions were hence selected to model severity distributions.

### 6.1.3 Monte Carlo Simulation

The Monte Carlo simulation used a frequency distribution to generate a random number representing the number of incidents which occurs during the upcoming year. The Severity distribution simulates what direct cost a bank could experience given that an incident occurs. The sum of each simulated incident during a year corresponds to the total cost a bank will face during an upcoming year, which is referred to the total aggregated loss. The number of trials was 10 000. There are two different frequency distributions and four different severity distributions that were relevant to test and all the frequency distributions have been tested with all of the severity distributions. This generated eight different aggregate loss distributions.

*Table 3 - Model's output from Method One*

| Model | Frequency distribution | Severity Distribution | Median | $VaR_{95\%}$ |
|---|---|---|---|---|
| **1** | Poisson | Pearson 5 | 0,525 | 1,16 |
| **2** | Poisson | Log Logistic | 0,369 | 0,548 |
| **3** | Poisson | Inverse Gaussian | 1,047 | 1,344 |
| **4** | Poisson | Lognormal | 0,428 | 0,519 |
| **5** | Negative Binominal | Pearson 5 | 0,525 | 1,219 |
| **6** | Negative Binominal | Log Logistic | 0,366 | 0,648 |
| **7** | Negative Binominal | Inverse Gaussian | 1,029 | 1,645 |
| **8** | Negative Binominal | Lognormal | 0,42 | 0,656 |

The numbers are presented in Billions of Euros. However, please keep in mind that the data these numbers are based on have been multiplied with a secret factor in order to keep the data anonyms.

### 6.1.4 Aggregate Loss Distribution

By repeating the simulated yearly loss output of each model in a Monte Carlo simulation 10 000 times, enough data have been gathered to form an aggregate loss distribution which shows the possible outcome and their corresponding probability. These distributions have in turn been interpreted and the best fitted continuous distribution is displayed in the table below. A graphical visualization is displayed in forms of bar charts and can be found in the appendix under Appendix 1.

*Table 4 - Best Fitted Aggregate Loss Distribution in Method One*

| Model | Best fitted aggregate loss distribution | AIC | BIC |
|---|---|---|---|
| **1** | Log-Logistic | 416 680 | 416 702 |
| **2** | Log-Logistic | 404 563 | 404 585 |
| **3** | Gamma | 421 842 | 421 863 |
| **4** | Lognormal | 416 740 | 416 762 |
| **5** | Log-Logistic | 416 783 | 416 805 |
| **6** | Log-Logistic | 416 838 | 416 860 |
| **7** | Gamma | 416 886 | 416 908 |
| **8** | Gamma | 416 790 | 416 812 |

## 6.2 Method Two

In Method two, the same frequency distributions are assumed that were used in Method one. The severity distributions, on the other hand, are modeled by a separate body distribution and a right tail distribution.

### 6.2.1 Severity Distribution

The body of the historical data were again best fitted by the same four distributions used for modelling severity in method one. However, the parameters of these distributions were calibrated differently and resulted in a lower AIC and BIC score. Notice that both the Inverse Gaussian distribution and the Lognormal distribution now make a better-fitted distribution then the Log-logistic distribution which previously was the second best-fitted distribution.

*Table 5 - The Body-Severity Distributions AIC and BIC Scores in Method Two*

| Continuous Body Distribution: | AIC | BIC |
|---|---|---|
| **Pearson 5** | 100 268 | 100 268 |
| **Inverse Gaussian** | 100 407 | 100 426 |
| **Lognormal** | 100 486 | 100 504 |
| **Log-Logistic** | 100 491 | 100 509 |

The tail of the historical data was best fitted by the Pareto distribution or the Inverse Gaussian distribution. The Pareto distribution achieved a better BIC score while the Inverse Gaussian obtained a better AIC score. The Exponential distribution were the third best fit but for simplicity reasons were not tested in this paper. The Exponential distribution was also believed to generate more extreme tail losses than actually occurs in reality. This is because the Exponential distribution generated a fatter tail than the data would suggest.

*Table 6 - The Tail-Severity Distributions AIC and BIC Scores in Method Two*

| Continuous Tail Distribution: | AIC | BIC |
|---|---|---|
| **Pareto** | 7 156 | 7 163 |
| **Inverse Gaussian** | 7 155 | 7 165 |

The purpose of Method two is to create a "hybrid" severity distribution which consists of a body and a tail distribution. Merging these body and tail distributions in all possible combination creates eight possible hybrid distributions which all have lower AIC and BIC scores compared to the severity distributions in method one.

*Table 7 - Hybrid-Severity Distribution used in Method Two*

| Hybrid | Body distribution | Tail Distribution | AIC | BIC |
|---|---|---|---|---|
| **1** | Pearson 5 | Pareto | 107 424 | 107 431 |
| **2** | Inverse Gaussian | Pareto | 107 563 | 107 589 |
| **3** | Lognormal | Pareto | 107 642 | 107 667 |
| **4** | Log-Logistic | Pareto | 107 647 | 107 672 |
| **5** | Pearson 5 | Inverse Gaussian | 107 423 | 107 433 |
| **6** | Inverse Gaussian | Inverse Gaussian | 107 562 | 107 591 |
| **7** | Lognormal | Inverse Gaussian | 107 641 | 107 669 |
| **8** | Log-Logistic | Inverse Gaussian | 107 646 | 107 674 |

### 6.2.2 Monte Carlo Simulation

The number of trials used in this Monte Carlo simulation is again 10 000. There are two different frequency distributions and eight different hybrid-severity distributions tried in these Monte Carlo simulations. All combinations of distributions are tested which means that there are 16 models a Monte Carlo simulations were performed on. Each combination of frequency and hybrid-severity distribution is referred to as an individual model (1-16) and correspond to an individual Monte Carlo simulation.

*Table 8 - Model's output from Method Two*

| Model | Frequency distribution | Hybrid Severity Distribution | Median | $VaR_{95\%}$ |
|---|---|---|---|---|
| **1** | Poisson | Hybrid 1 | 0,702 | 3,438 |
| **2** | Poisson | Hybrid 2 | 0,685 | 3,445 |
| **3** | Poisson | Hybrid 3 | 0,674 | 3,643 |
| **4** | Poisson | Hybrid 4 | 0,678 | 3,556 |
| **5** | Poisson | Hybrid 5 | 0,726 | 1,747 |
| **6** | Poisson | Hybrid 6 | 0,707 | 1,708 |
| **7** | Poisson | Hybrid 7 | 0,693 | 1,749 |
| **8** | Poisson | Hybrid 8 | 0,706 | 1,713 |
| **9** | Negative Binominal | Hybrid 1 | 0,702 | 3,438 |
| **10** | Negative Binominal | Hybrid 2 | 0,685 | 3,445 |
| **11** | Negative Binominal | Hybrid 3 | 0,674 | 3,643 |
| **12** | Negative Binominal | Hybrid 4 | 0,678 | 3,556 |
| **13** | Negative Binominal | Hybrid 5 | 0,726 | 1,747 |
| **14** | Negative Binominal | Hybrid 6 | 0,707 | 1,708 |
| **15** | Negative Binominal | Hybrid 7 | 0,693 | 1,749 |
| **16** | Negative Binominal | Hybrid 8 | 0,706 | 1,713 |

The numbers are presented in Billions of Euros. However, please keep in mind that the data these numbers are based on have been multiplied with a secret factor in order to keep the data anonyms.

### 6.2.3 Aggregate Loss Distribution

Similar to Method one, these aggregated loss distributions obtained from the Monte Carlo simulations have, in turn, been interpreted and are visually displayed in Appendix 1 in forms of bar charts. Below is a table of the best-fitted distribution for the aggregated loss output, suggested by the software.

*Table 9 - Best Fitted Aggregate Loss Distribution in Method Two*

| Model | Best fitted aggregate loss distribution | AIC | BIC |
|---|---|---|---|
| **1** | Log-Logistic | 428 754 | 428 775 |
| **2** | Log-Logistic | 428 612 | 428 634 |
| **3** | Log-Logistic | 428 904 | 428 926 |
| **4** | Pearson 5 | 428 027 | 428 048 |
| **5** | Lognormal | 420 692 | 420 714 |
| **6** | Inverse Gaussian | 420 302 | 420 323 |
| **7** | Lognormal | 420 608 | 420 630 |
| **8** | Lognormal | 420 555 | 420 577 |
| **9** | Log-Logistic | 432 103 | 432 124 |
| **10** | Log-Logistic | 432 504 | 432 525 |
| **11** | Pearson 5 | 431 854 | 431 875 |
| **12** | Log-Logistic | 432 160 | 432 181 |
| **13** | Lognormal | 424 411 | 424 432 |
| **14** | Lognormal | 424 194 | 424 216 |
| **15** | Lognormal | 424 012 | 424 034 |
| **16** | Lognormal | 424 093 | 424 114 |

# 7. Analysis

*The purpose of this chapter is to present the analysis around the presented result obtained from the models and to give a discussion about the models accuracy and performance.*

## 7.1 Method One

### 7.1.1 Frequency Distribution

The Poisson distribution and Negative Binominal distribution are chosen to model the frequency because of their good AIC and BIC scores. This paper chose to ignore any other discrete distribution like the geometric distribution to model the frequency because of the significant difference in AIC scores. The uniform distribution were not used since the frequency of the data were clearly not uniformly distributed. Because all other discrete distributions scored such a poor AIC and BIC score, including these distributions in a model would mean a great loss of information from the original data.

### 7.1.2 Severity Distribution

The severity distribution were a lot harder to model since the distributions which showed the lowest AIC and BIC scores were still not as close to the actual data as the frequency distribution was. Part of the very high AIC and BIC scores can be explained by the high number of data points included in the calculations. However, visual interpretations of the suggested distributions and the actual data show that the best-fitted severity distribution is still not relatively close to reality. Although these four distributions were the ones which came closest, with the Pearson 5 distribution barely in the lead.

For these four distributions, the AIC and BIC scores were relative close to each other. However, other distributions were also suggested but these AIC and BIC scores were not relatively close to the best four, so these distributions were ignored.

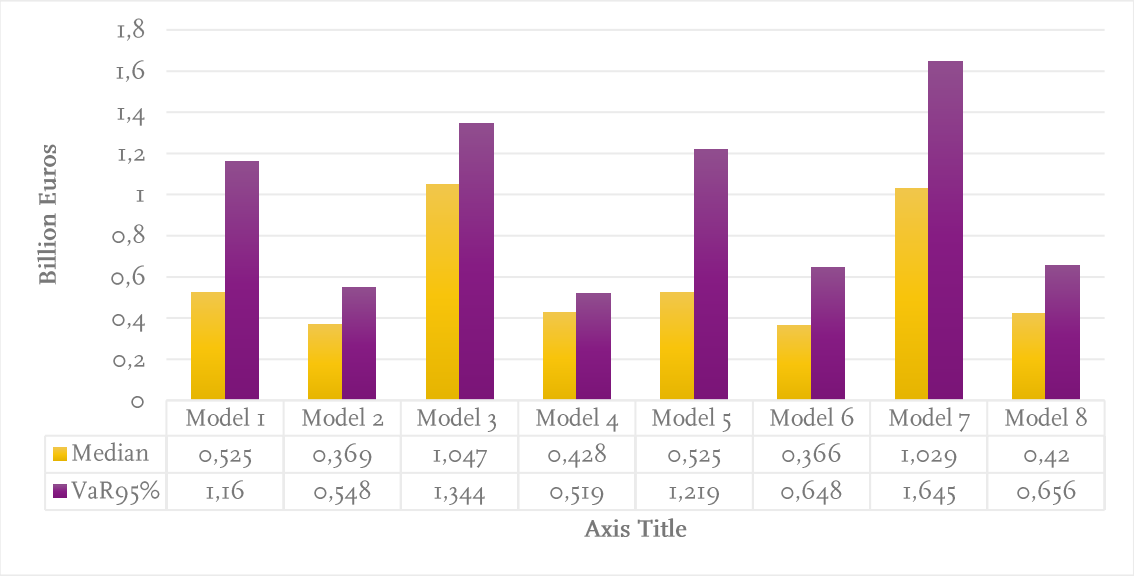
### 7.1.3 Monte Carlo Simulation

The Monte Carlo simulation on each of the eight models in method one showed a varied range of result. Most model's aggregate loss distribution are skewed which proved to be a very common result for all the simulations, including method two. This is not a surprising outcome since the external data is characterized with this skew.

It turned out that the aggregate loss distribution was very hard to fit a distribution to because of its extreme values of its outliers. Assuming that the aggregated yearly loss of IT-incidents follows a specific distribution is therefore not a good assumption to make. It is better looking at the median value and a VaR value of the data output from the Monte Carlo simulation when making a prediction of future losses. The 95 % VaR level are presented in the result but it is important to also look at the further extreme values and the density of these extreme values when evaluating the models. A graphical representation can be seen in the appendix under Appendix 1.

Below is a bar chart displaying a graphical representation of model 1 to 8 Median and $VaR_{95\%}$ value presented in table 3. The median value of each model is represented by the yellow bar while the purple bar represents the 95 % VaR value.

*Figure 1 - Graphical Comparison of the Models in Method One*




| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 | Model 7 | Model 8 |
|---|---|---|---|---|---|---|---|---|
| Median | 0,525 | 0,369 | 1,047 | 0,428 | 0,525 | 0,366 | 1,029 | 0,42 |
| VaR95% | 1,16 | 0,548 | 1,344 | 0,519 | 1,219 | 0,648 | 1,645 | 0,656 |

When looking at the simulated output from each model we can observe that all models except 3 and 7 showed very similar result in both their median value and VaR value. (See figure 1). Model 3 and 7 showed the result which was a lot bigger than the rest, both in terms of median and VaR. These models generated an aggregated distribution which was located at higher values. Model 3 and 7 do have one thing in common though, they both assumed the Inverse Gaussian distribution as the severity distribution.

Model 3 and 7 showed the results which were most accurate to the external data in terms of median and VaR. Closest are model 7, with both of median and VaR closest to the median and VaR of the external data. Model 7 did also agree with the external data the best through a visual comparison when comparing the body of the distributions. However, even as this model seemed to be the best of these at explaining the aggregate loss distribution out of these 8, it does a poor job at representing an accurate right tail. In fact, all eight models fail to sufficient capture to tail distribution.

An observation made here are the similarities in both median and VaR values of model 2 and 4. The same goes for model 6 and 8. Model 2 and 6 both assumes log-logistic severity distribution while the models 4 and 8 assumes lognormal severity distribution. These two models seems to produce similar result, only differ slightly because of the different frequency distribution assumed. This is interesting since it signals that the log-logistic distribution and lognormal distribution are, with this calibration, very similar in their characteristics and could potentially be a good alternative to each other.

## 7.2 Method Two

### 7.2.1 Frequency Distribution

Method two makes no adjustments to the frequency distributions. The discrete frequency distribution was easy to model and fit very well with the actual data, both from visual interpretation as well as interpreting the AIC and BIC scores.

However, because of the changing digital environment, it could be argued if a drift component should be added to the frequency assumptions in the model. The problem is at what direction the drift component should be directed at. Should it increase the number of yearly IT incidents because of the stronger dependency on IT? Or should it decrease the number of incidents because of the banks quickly improving their IT systems and becoming more efficient in managing IT failure? It is most likely that the number of yearly IT incidents will go down in the future which would argue for a negative drift component. However, no negative trend have been concluded in this paper so a drift component have not included in the models but leaves this door open for future improvements.

### 7.2.2 Severity Distribution

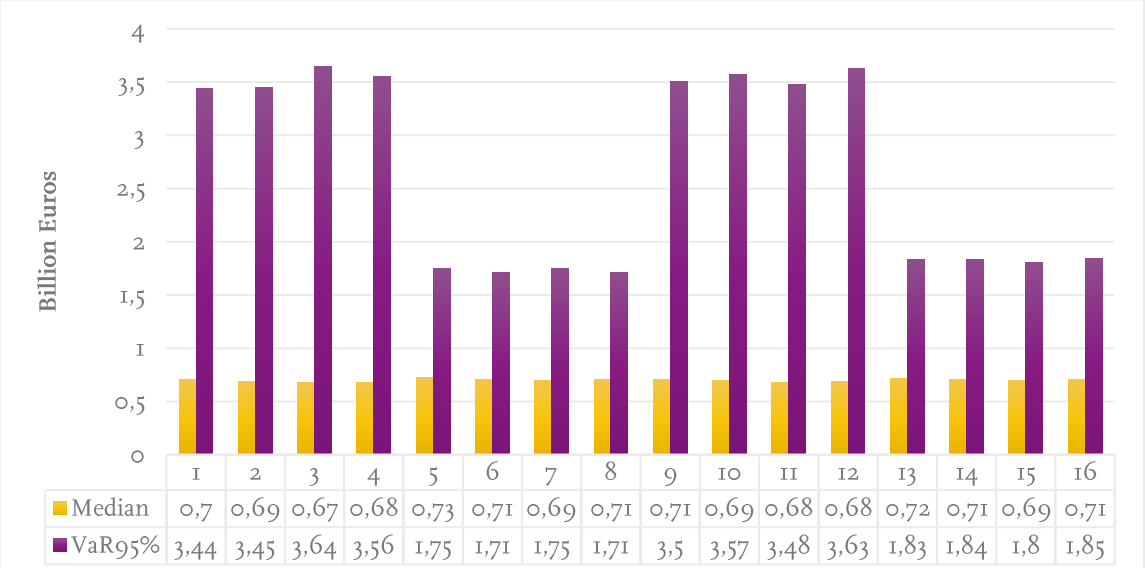
When modelling the body distribution is that the same continuous distributing were suggested as before. These are the Pearson 5, Log-logistic, Inverse Gaussian and the Lognormal distribution. A noticeable difference is the order in which these distributions were suggested. The log-logistic distribution went from being the second best fit in method one to be the fourth best fit of the body distribution in method two. Not surprisingly since the extreme data that made the log-logistic distribution a good fit were now modeled by the tail distribution instead.

All of the hybrid alternatives for the severity distribution used in method two resulted in a better fit than the severity distributions used in method one. This could both be seen from the aggregated AIC and BIC which were lower for method two, and from visual interpretation. A hybrid alternative to model severity would, therefore, be a good approach to take when modelling IT risk, or other operational risks with the same type of extreme outliers.

### 7.2.3 Monte Carlo Simulation

Below is a bar chart displaying a graphical representation of model 1 to 16's Median and $VaR_{95\%}$ value presented in table 8. The median value of each model is represented by the yellow bar while the purple bar represents the 95 % VaR value, all values are presented in billions of Euros.

*Figure 2 - Graphical Comparison of the Models in Method Two*




| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Median | 0,7 | 0,69 | 0,67 | 0,68 | 0,73 | 0,71 | 0,69 | 0,71 | 0,71 | 0,69 | 0,68 | 0,68 | 0,72 | 0,71 | 0,69 | 0,71 |
| VaR95% | 3,44 | 3,45 | 3,64 | 3,56 | 1,75 | 1,71 | 1,75 | 1,71 | 3,5 | 3,57 | 3,48 | 3,63 | 1,83 | 1,84 | 1,8 | 1,85 |

The Monte Carlo simulation on each of the 16 models produced median values very close to each other (see figure 2). The smallest median is generated from model 3 and 11 ended up around 0,674. The biggest estimated median is generated from model 5 and 13 and landed around 0,726. So it was a really small gap relatively speaking. Interestingly, both model 3 and 11 had the same tail and body distribution assumptions. The same observation can be made for model 5 and 13 as well. This leads to the conclusion that the body and tail distributions are the most pivotal distributions assumptions in a LDA model and will affect the result the most. The frequency distribution does not make any major impact on the result at all in this method. There are small variations between models in method one which makes different frequency distribution assumptions. But these variations are small and only affects the VaR. It cannot be excluded that these variations are the result of the randomness involved in a Monte Carlo simulation. To significantly disregard this randomness effect, the Monte Carlo simulations should be using a higher number of trials, preferably 100 000 times or more.

It was equally difficult to fit a continuous distribution to the aggregate loss distribution in this method as it was in method one. Instead, it is more informative looking at VaR value of the data output from the Monte Carlo simulation when making a prediction of future losses and comparing with the original historical data. See appendix 1 for a visual representation of aggregated loss distributions.

All models in Method two generated very accurate body distributions. Where the median is very close to the external data, as well as a good agreement between the resulting body distribution and the body distribution of the external data. The models differ in the extreme values however. Models 5-8 and 13-16 underestimate the extreme values that the external data can reach and are, therefore, bad models to use. Better models to use are therefore model 1-4 and 9-12. What these models have in common is that they all use the Pareto distribution as a tail distribution assumption. The Pareto has therefore proven to be the best distribution for modelling tails which includes very extreme outliers, like the IT risk. This is also interesting since it is the same conclusion that Jöhnemark (2012) made in his research. This result also shows the importance of getting the right tail distribution. It is essential to make the right distribution assumptions in general but the tail distribution is the most important one for this risk (and probably for other operational risks with similar characteristics).

When looking at the simulated output from each model we can observe that the mean value was really high, sometimes as high as the 90 % VaR level. This is because of the extreme outliers which move the mean. This is the reason why the median is a better predictor of the expected value for this risk. Furthermore, a very high mean can also be observed in the external data which leads to the conclusion that the models are correct. This being said, some models have proven to generate so extreme values they do not make much intuitive sense.

### 7.2.4 Best model for IT risk

The most accurate model of them all in Method two was model 4. This model showed the realistic result for the body as well as made a really good prediction of VaR values that correspond well with the external data. This model also shows simulated losses that are far bigger than ever occurred before during these past 10 years. But since this model proves to show a great agreement with the external data, it can be argued that these losses are possible nevertheless. This is very interesting and should be something to consider for a risk manager working with the IT risk. Model 4 showed a median of 0,687 Billion euros and a $VaR_{95\%}$ of 3,556 Billion. The $VaR_{99\%}$ landed on 16,963 billion instead, which was never reached by the actual historical yearly losses computed from the external data. But because these are just 10 years to compare with, it cannot be concluded that the worst case scenario is observed within this historical data. In a really bad year where most incidents cause major losses, the 99 percentile produced by the model could be plausible and is something to look out for. These extreme events are proven to be really rare (1 out of 100 for the whole industry) but if they were to occur it would definitely put some affected banks in a difficult situation. This paper also only considers the direct cost made by these incidents. The indirect cost could actually cause just as much damage, or even more. The public's trust in a bank could be seen as its biggest asset, which could be severely damaged by major IT incidents like cyberattacks. It is therefore well worth to keep an eye on the IT risk.

The ideal model did, therefore, consist of a Poisson frequency distribution, Log-Logistic severity-body distribution, and a Pareto severity-tail distribution.

## 7.3 Discussion

### 7.3.1 Distribution Assumptions

An interesting observation made in this work was that the Poisson did not appear to be the closest fit for the frequency distribution looking at the AIC and BIC scores. However, it did make intuitive sense assuming this distribution when making a visual comparison and when considering the general characteristics of the frequency. Although the frequency distribution made did not have major impacts on the models, the best model to estimate future losses (Method two, Model 4) did have the Poisson frequency distribution assumption. This is an interesting result since Poisson distribution was suggested to be the best discrete distribution to model the frequency of operational risk in the work made by Svensson (2015).

The Severity distribution that is assumed in Method one and as body distribution in Method two were all very similar in their characteristics. For Method one the Inverse Gaussian distribution did the best job in as a severity distribution of IT risk. It makes sense because this distribution when calibrated to the whole data gave the fattest tail for the very extreme values which gave the aggregated loss distribution a more accurate tail than the rest. However, the models in Method one did not produce an accurate result which makes Method one not a good method to use in modelling this IT risk. Method two was a lot better to use instead were the best model used a Log-logistic distribution as body distribution. This distribution is therefore considered a better representative of only the body data when calibrated correctly. All in all, the Pearson 5, Log-logistic, Lognormal and Inverse Gaussian distribution are all very similar to each other in their positive skew. This skew represents the characteristics of the main data very well. However, the distributions differ when it comes to replicating the right tail which no one did a good job at. Hence, they should be used for modelling the main body of the data while assigning a separate continuous distribution for the remaining extreme values, possibly the Pareto distribution.

### 7.3.2 Quality of the Model

When the historical data is analyzed, It looked like a downwards trend is present which makes intuitive sense considering that IT systems are only getting more efficient and reliable. The big amount of money being invested in digitalization of the financial industry seems to be leading to fewer losses caused by IT incidents. So even though more and more processes within a financial institution are being digitalized and more dependent on IT systems, the IT

systems gets more reliable which in the end decreases the yearly losses for the industry. This is an observation made strictly over the industry and the situation could just as well be different for the individual institutions. The models would therefore perhaps be improved by adding a drift component to the frequency distribution assumption. This paper has not included a drift component but leaves this question as a suggestion for further research.

A hybrid severity distribution is a better approach to model IT risk and possibly other operational risks that has the same characteristics in their data. Recent history has shown that the banking industry has a somewhat special exposure to operational risk where incidents could be so extreme it could be the primary threat for a bank. Examples can range from legal disputes, fraud scandals to hacker attacks. It is important for the risk managing department to know what type of losses these rare events could inflict in a worst case scenario and the hybrid models do a good job at estimating these events.

### 7.3.3 Merger of Body and Tail Distributions

The severity distribution in Method two is modeled by a hybrid distribution. This hybrid distribution separates the extreme-valued losses from the more common valued losses by having a right-tail and a body distribution separately modeled and then merged. The merger of this two distribution was done in a method inspired by Fabio Piacenza's work: R and Operational Risk. It is possible that there could be a better way of merging distributions which have not been covered by the literature which this paper is based on. Perhaps calibrating the distribution parameters in order to normalize the simulated output slightly. By adjusting the two distributions parameters, it is possible to get these two distributions closer to each other and perhaps achieve a better merger. However, it is unclear if this would generate a more accurate result. The two distributions are already calibrated to the actual data were one is calibrated to the body data and the other one to the extreme data. Adjusting this calibration could result in a worse fit and therefore decrease the accuracy of the result. On the other hand, constructing a hybrid severity distribution like this paper has done, could lead to a polarizing of the estimated losses. Were many of the simulated losses have a value close to the median and a few losses have an extremely high value. But at the same time, this is what we observe when looking at the actual historical data for IT risk, leading to the conclusion that this paper used a well-functioning method for merging the body and tail distributions. This question has not been the focus of this paper and is therefore left as a suggestion for further research.

### 7.3.4 Limitation to the Thesis

The quantification made in this paper is done on low frequency high direct cost data. Meaning it is only extreme-valued events that are included. The most common IT incidents do not realize any direct cost for the financial institution or such a low cost that it most of the times are insignificant for the institution. This limits the application of the model suggested to only predict these types of incidents as well.

The suggested model from this paper is the model that have been the most accurate to the external data and generated the most feasible result given the characteristics of this risk. The model can only be considered best for quantifying IT risk in the banking industry and can only be considered the best model out of the 24 models tested in this paper.

The main purpose of this paper was to find the best model for this type of risk and explaining the process of how to get this model. Even though the data used in this paper was multiplied by a secret factor for confidentiality reasons this objective has not been affected. Since all data points were multiplied by the same factor it does not affect the relative distribution of the external data, only the location parameters. This means that the same distribution assumptions will be made regardless. However, the calibration of these distributions could differ which would, in that case, affect what body-severity distribution to assume.

### 7.3.5 Suggestion for Further Research

It would be interesting to see if a declining trend could be observed in the number of IT incidents occurring in the banking industry. If a trend of declining frequency of IT risk can be significantly proven, then a drift component could be added to the model. A suggestion on further research is therefore to seek out this trend and, if its existence is proven, to recommend a reasonable drift component to add to this model.

An academic research exploring the best way to properly merge a tail and body distribution in hybrid severity distribution could not be found. It is still very interesting for this paper to know what other alternative merging methods could be used and show their advantages and disadvantages. This is therefore another suggestion for further research.

# 8. Conclusion

*This final chapter will provide a summary of the findings and made by the work presented in this paper in order to answer the original question.*

The appropriate way to modelling IT risk for the banking industry is to use Model 4 from Method two.

In general terms, Method two is the best method to use for modelling the IT risk. Hybrid models do a better job at estimating rare events with high severity and is, therefore, a good method to use for quantifying the IT risk. Model 4 from method two is the best model to use out of the 24 models tested in this paper, for quantifying the IT risk exposure in the banking industry.

Frequency distribution assumption has little impact on the result but Poisson distribution is the slightly better distribution to use. The severity distribution assumption was the most important assumption to make and the best one to use depends on the method used for modelling severity. In Method one, the Inverse Gaussian distribution was the best distribution to use. In Method two, the tail distribution assumption was the most pivotal where Pareto tail distribution proved to be the best. The body distribution did not affect as much but still had an impact and the Log-logistic distribution was the best one to use.

## 9. References


Acharyya, D. M. (2012). *Why the current practice of operational risk management in insurance is fundamentally flawed - evidence from the field*. Bournemouth: Bournemouth University Business School.

Andale. (2017, 05 13). *Statistics How To*. Retrieved from www.Statisticshowto.com: http://www.statisticshowto.com/

Aue and Kalkbrener. (2007). *LDA at Work*. Frankfurt: Deutsche Bank.

Aven, T. (2003). *Foundations of Risk Analysis*. Chichester: John Wiley & Sons Ltd.

Bakker, M. (2004). *Quantifying Operational Risk within Banks according to Basel II*. Delft: Delft Institute of Applied Mathematics (in cooperation with PwC).

Bank for International Settlements. (2004, July 01). *Implementation of Basel II: Practical Considerations*. Retrieved from ww.BIS.org: http://www.bis.org/publ/bcbs109.htm

BIS. (2016). *Standardised Measurement Approach for operational risk - consultative document*. Basel: Bank for International Settlements.

Broeders and Khanna. (2015). Strategic choices for banks in the digital age. *McKinsey&Company*, 1.

Clark and Thayer. (2004). *A Primer on the Exponential Family of Distribution*. Arlington: Casualty Actuarial Society.

Degen Embrechts and Lambrigger. (2007). *The Quantitative Modeling of Operational Risk Between G-and-H EVT*. Cambridge: Cambridge University.

EBA. (2017, 05 09). *Operational risk*. Retrieved from eba.europa.eu: https://www.eba.europa.eu/regulation-and-policy/operational-risk

ECB. (2016). *Stocktake of IT risk supervision practices* . Frankfurt: European Central Bank.

Fontnouvelle Rueff Jordan and Rosengren. (2003). *Using Loss Data to Quantify Operational Risk*. Boston: Federal Reserve Bank of Boston.

Frachot and Roncalli. (2002). *Mixing internal and external data for managing operational risk*. Lyon: Groupe De Recherche Opérationnelle.

Franchot Georges and Roncalli. (2001). *Loss Distribution Approach for operational risk*. Lyon: Groupe de Recherche Opérationnelle.

Frost, J. (2017). *The Exciting Guide To Probability Distributions - Part 1*. Oxford: University of Oxford.

Gill Stewart Treasure and Chadwick. (2008). Methods of data collection in qualitative research: interviews and focus groups. *BDJ*, 1.

Guillen Gustafsson Nielsen and Pritchard. (2007). *Using External Data in Operational Risk*. Barcelona, Copenhagen, London: The Geneva Papers / The International Association for Study of Insurance Economics.

Hamedani, C. A. (2000). *THE BETA ODD LOG-LOGISTIC GENERALIZED FAMILY OF DISTRIBUTIONS*. Ankara: Hachettepe University.

Handbook Engineering Stastistics. (2017, 05 18). *Family of Distributions*. Retrieved from http://www.itl.nist.gov: http://www.itl.nist.gov/div898/handbook/eda/section3/eda363.htm

J.M. Hammersley and D.C. Handscomb. (1979). *Monte Carlo Methods*. New York: Chapman and Hall.

Jones, J. (2017, 05 18). *Stats: Probability Distrbution*. Retrieved from https://people.richland.edu/james/: https://people.richland.edu/james/lecture/m170/ch06-prb.html

Jöhnemark, A. (2012). *Modeling Operational Risk*. Stockholm: Royal Institute of Technology .

Keller and Bayraksan. (2011). Case-Quantifying Operational Risk in Financial Institutions. *Institute for Operations Research and the Management Science (INFORMS)*, 1-9.

Lahcene, B. (2013). *On Pearson families of distributions and its applications*. Ha'il: Department of Mathematics, Hail University.

Liddle, A. R. (2008). *Information criteria for astrophysical model selection*. Brighton, Honolulu: University of Sussex, University of Hawaii.

Mignola, G. (2003). *Integrating Internal and External data - Sanpaolo IMI perspective*. New York: Gruppo Sanpaolo IMI.

Navarrete, E. (2006). *Practical Calculation of Expected and Unexpected Losses in Operational Risk by Simulation Methods*. Ithaca: Palisade.

Piacenza, F. (2012). *R and Operational Risk*. Milano: UniCredit.

Power, M. (2003). *The Invention of Operational Risk*. London: ESRC Centre for Analysis of Risk and Regulation.

Ratner, C. (2002). Subjectivity and Objectivity in Qualitative Methodology. *Forum: Qualitative Social Research*, 1.

R-forge distribution Core Team. (2009). *A guide on probability distributions*. R-Forge.

Shevchenko, P. (2011). *Modelling Operational Risk Using Bayesian Inference*. Berlin: Springer-Verlag .

Standfordphd.com. (n.d.). *BAYESIAN INFORMATION CRITERION*. Retrieved from Statistical & Financial Consulting by Stanford PhD: http://www.stanfordphd.com/BIC.html

Svensson, K. P. (2015). *A Bayesian Approach to Modeling Operational Risk When Data is Scarce*. Lund: Lund University.

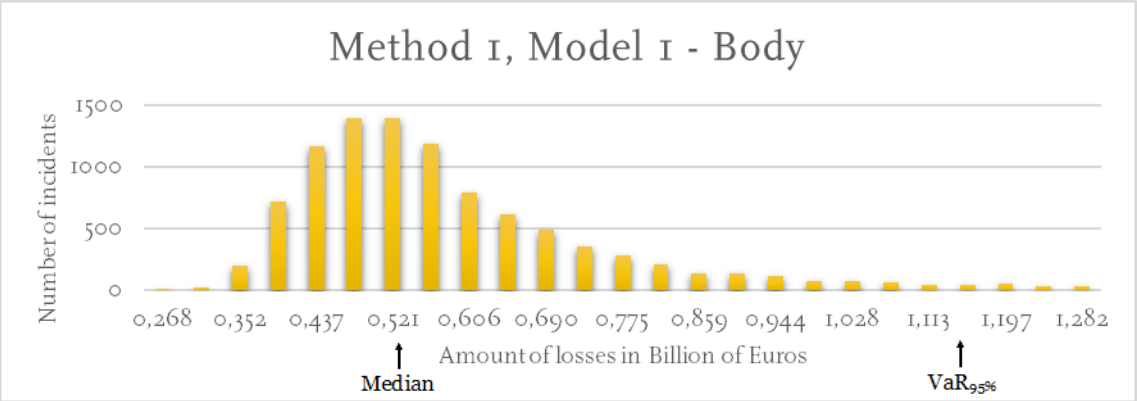
The Institute of Operational Risk. (2010). *Operational Risk Sound Practice Guidance*. Ware: The Institute of Operational Risk.

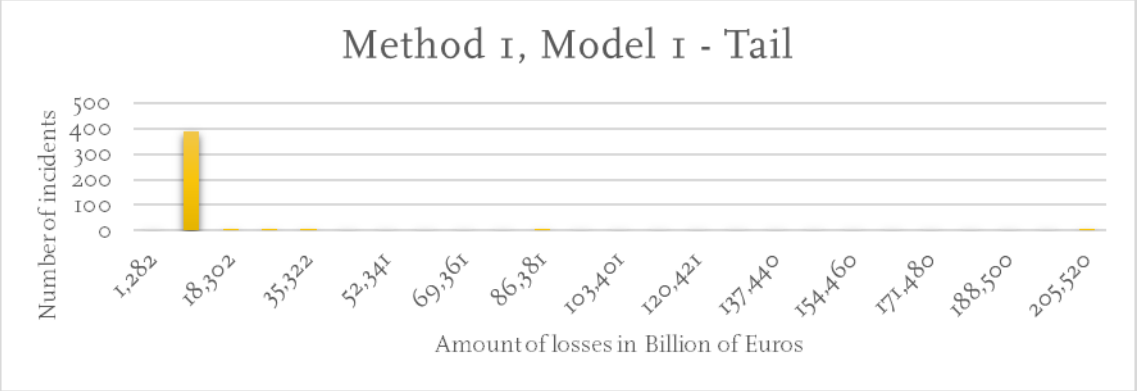
Thomas J. Linsmeier and Niel D. (n.d.). Value at Risk. *Pearson Financial Journal*.

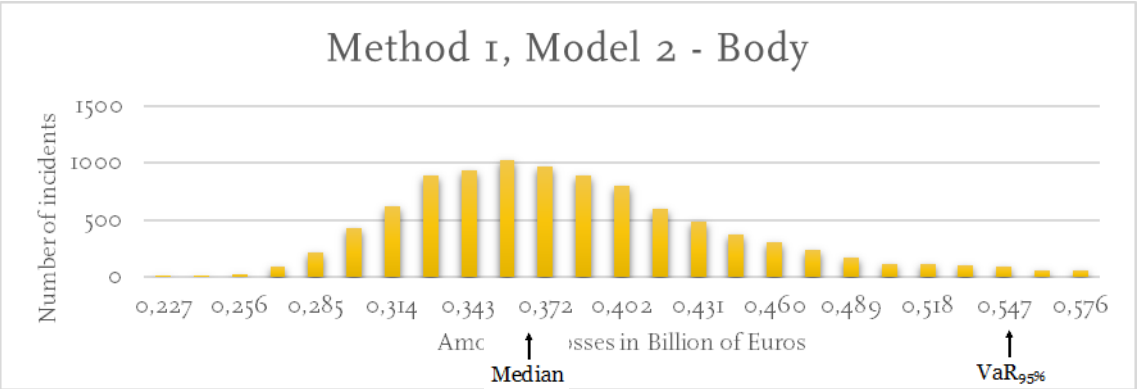
# Appendix

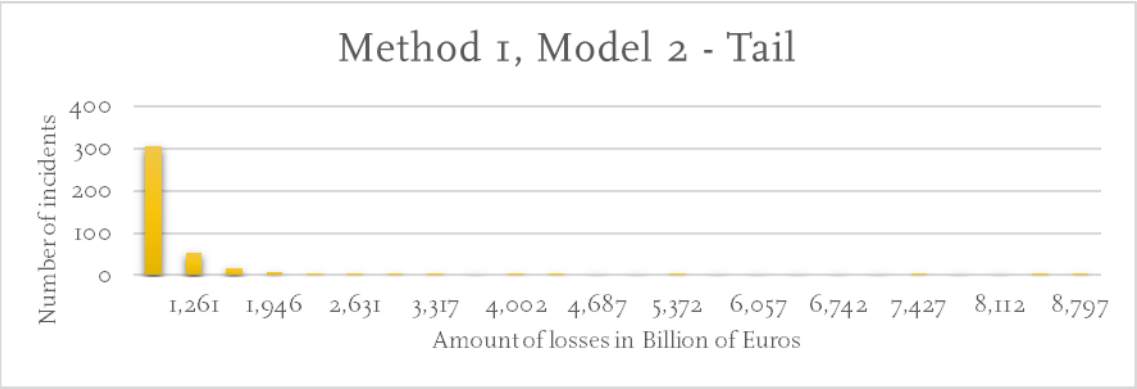
*Appendix 1 - Distribution of the Model's Output*

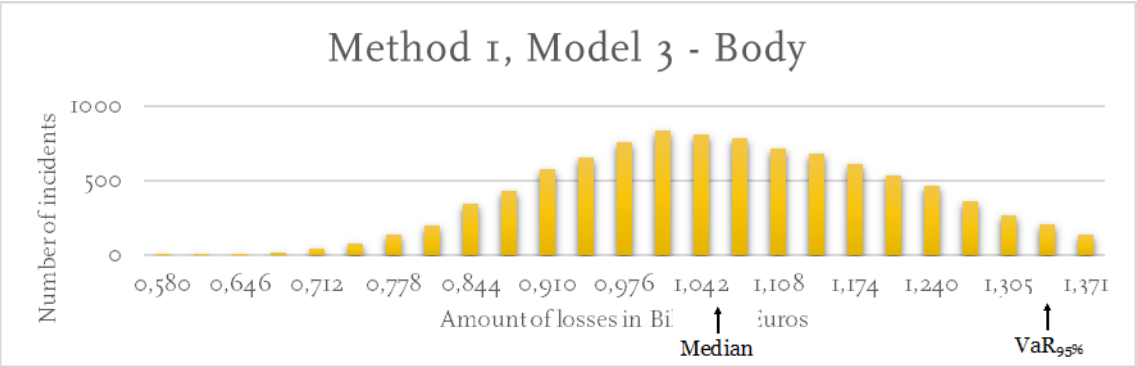
Following is the simulated result from the eight models used in method one and 16 models used in method two displayed in charts. Keep in mind that the tail, which goes on quite far beyond the body, are represented in an individual chart and show a lot bigger intervals which lead to the very large number of incidents in the start. They are in fact smoothly allocated in a diminishing manner from the body. The threshold between body and tail values in these graphs are for consistency again 96,05 %.

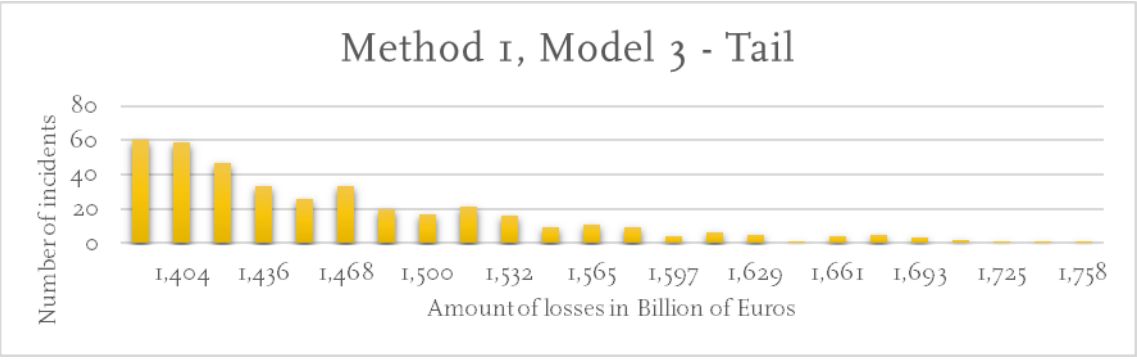


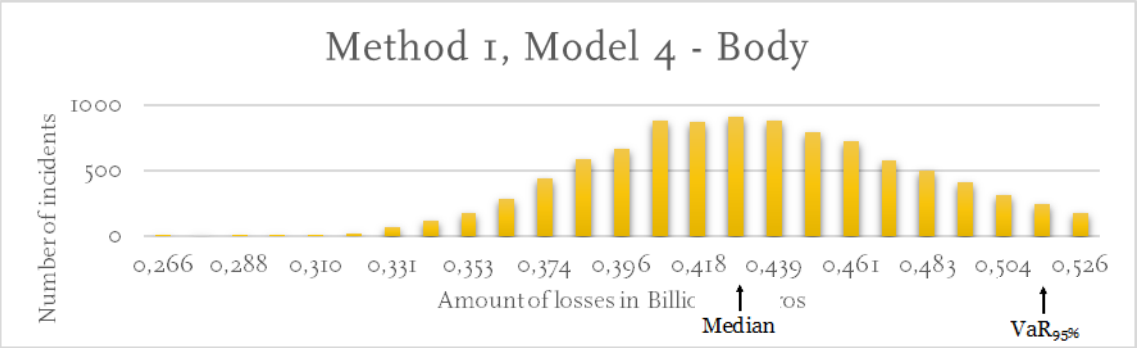



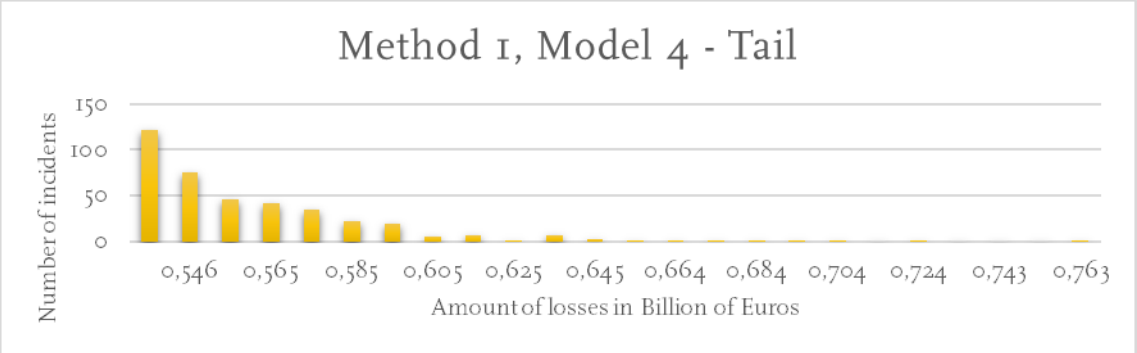


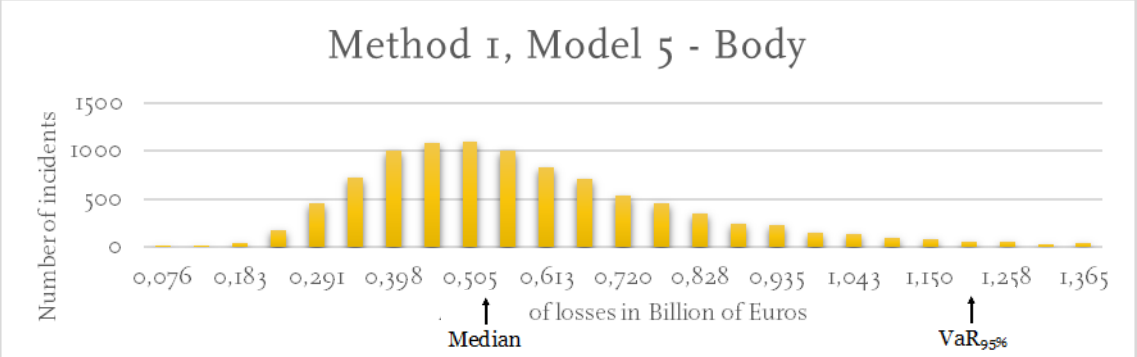



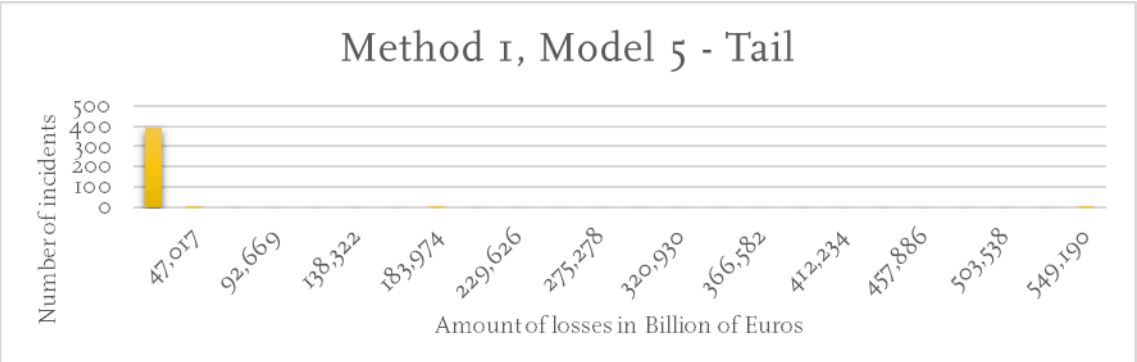



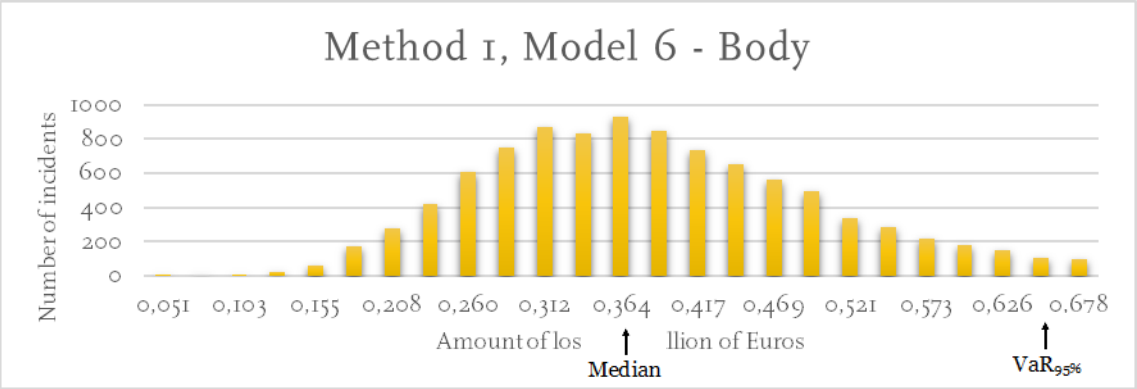



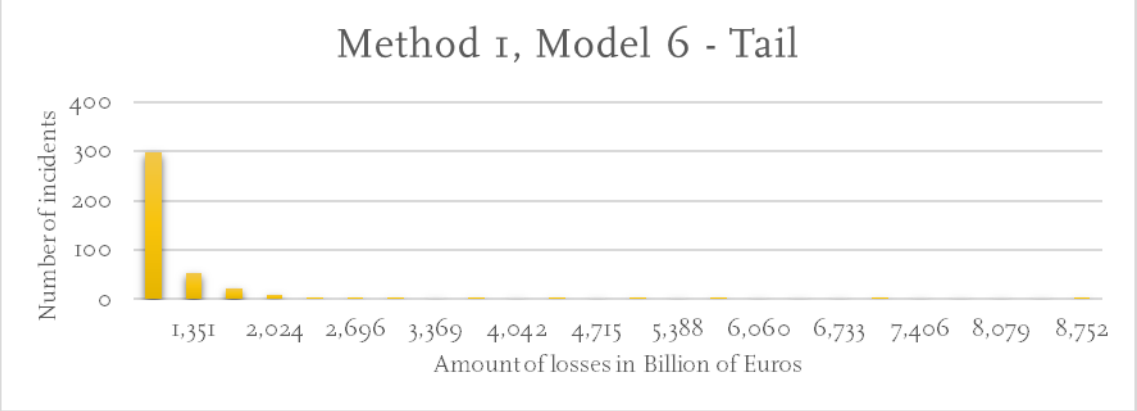

Method 1, Model 4 - Body
Number of incidents
1000
500
0
0,266 0,288 0,310 0,331 0,353 0,374 0,396 0,418 0,439 0,461 0,483 0,504 0,526
Median
VaR95%

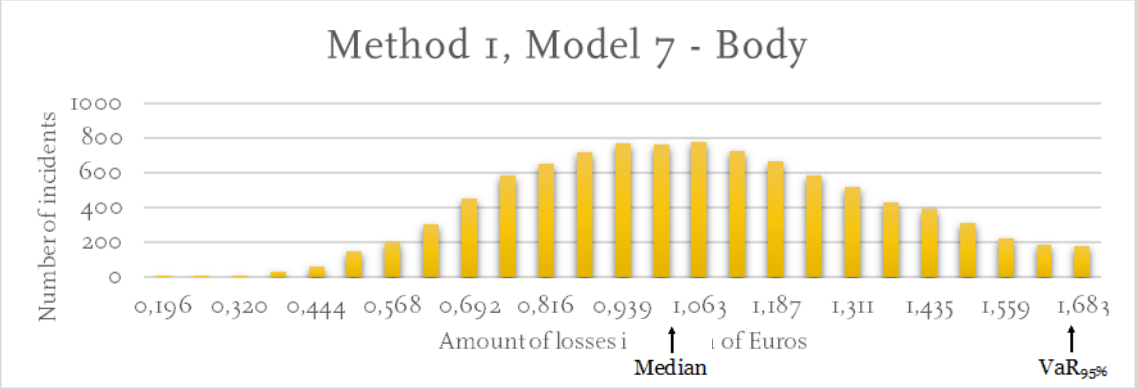

Method 1, Model 4 - Tail
Number of incidents
150
100
50
0
0,546 0,565 0,585 0,605 0,625 0,645 0,664 0,684 0,704 0,724 0,743 0,763
Amount of losses in Billion of Euros

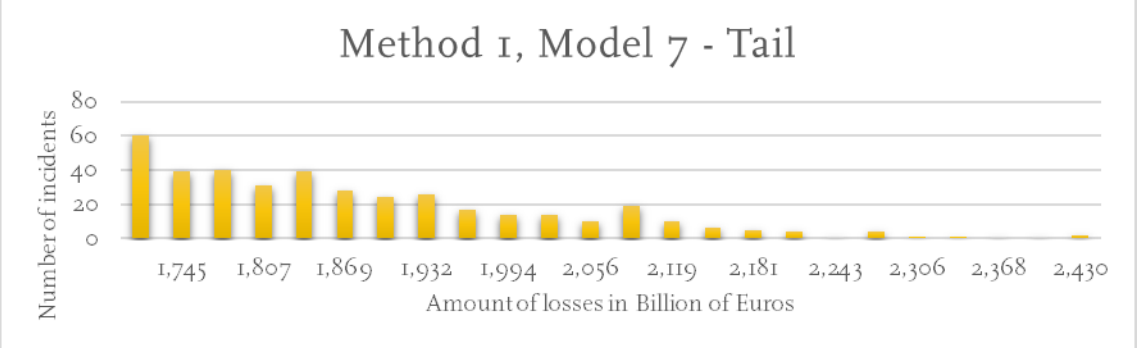

Method 1, Model 5 - Body
Number of incidents
1500
1000
500
0
0,076 0,183 0,291 0,398 0,505 0,613 0,720 0,828 0,935 1,043 1,150 1,258 1,365
of losses in Billion of Euros
Median
VaR95%

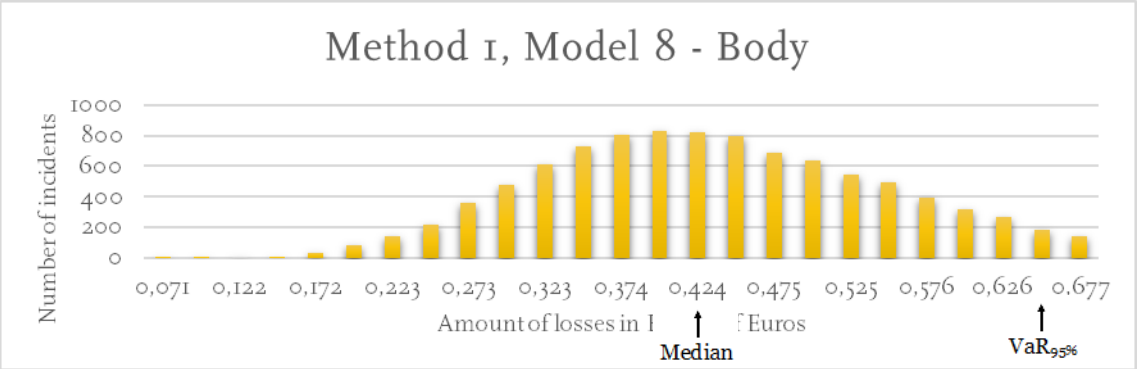

Method 1, Model 5 - Tail
Number of incidents
500
400
300
200
100
0
47,017 92,669 138,322 183,974 229,626 275,278 320,930 366,582 412,234 457,886 503,538 549,190
Amount of losses in Billion of Euros

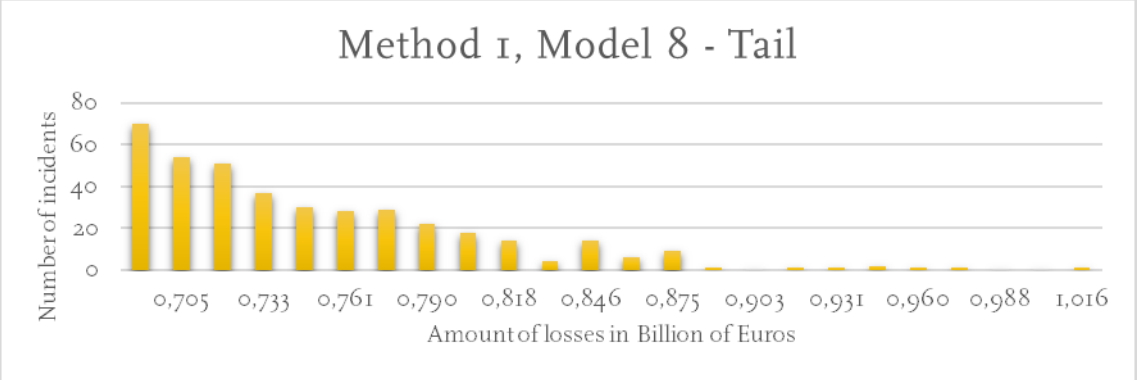

Method 1, Model 6 - Body
Number of incidents
1000
800
600
400
200
0
0,051 0,103 0,155 0,208 0,260 0,312 0,364 0,417 0,469 0,521 0,573 0,626 0,678
Amount of los llion of Euros
Median
VaR95%

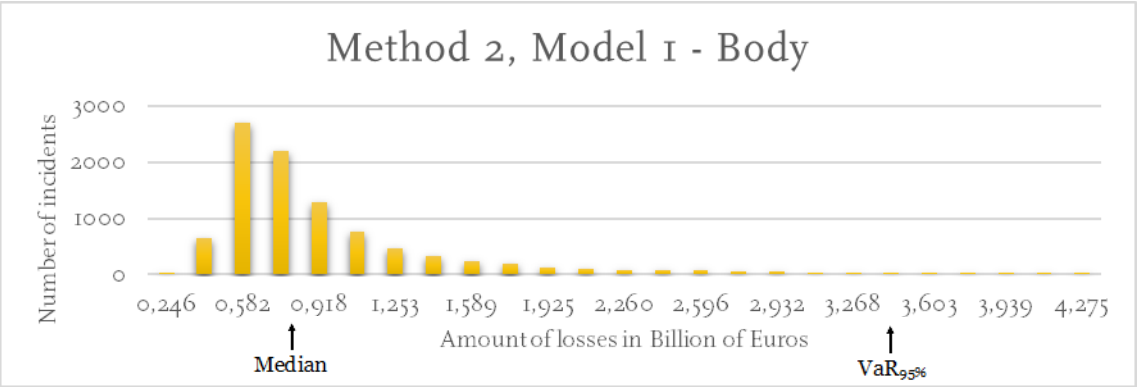

Method 1, Model 6 - Tail
Number of incidents
400
300
200
100
0
1,351 2,024 2,696 3,369 4,042 4,715 5,388 6,060 6,733 7,406 8,079 8,752
Amount of losses in Billion of Euros

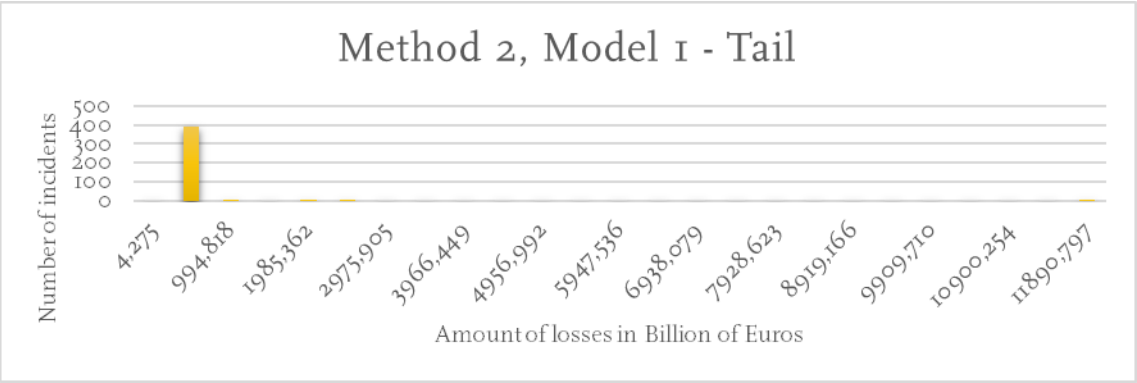

Method 1, Model 7 - Body
Number of incidents
1000
800
600
400
200
0
0,196 0,320 0,444 0,568 0,692 0,816 0,939 1,063 1,187 1,311 1,435 1,559 1,683
Amount of losses i of Euros
Median
VaR95%

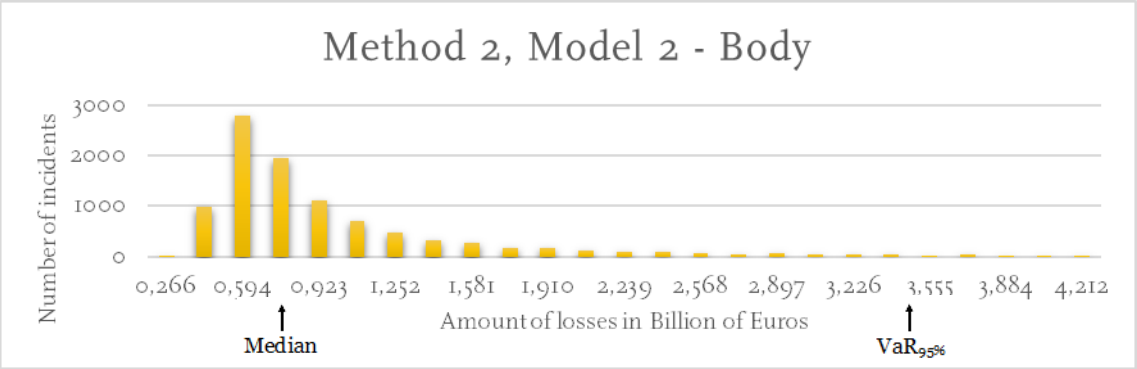

Method 1, Model 7 - Tail
Number of incidents
80
60
40
20
0
1,745 1,807 1,869 1,932 1,994 2,056 2,119 2,181 2,243 2,306 2,368 2,430
Amount of losses in Billion of Euros

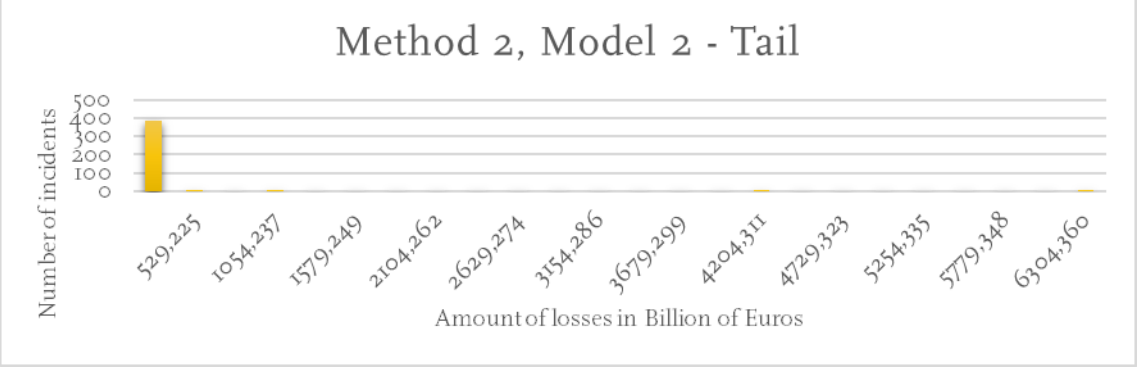

Method 1, Model 8 - Body
Number of incidents
1000
800
600
400
200
0
0,071
0,122
0,172
0,223
0,273
0,323
0,374
0,424
0,475
0,525
0,576
0,626
0,677
Amount of losses in I Euros
Median
VaR95%

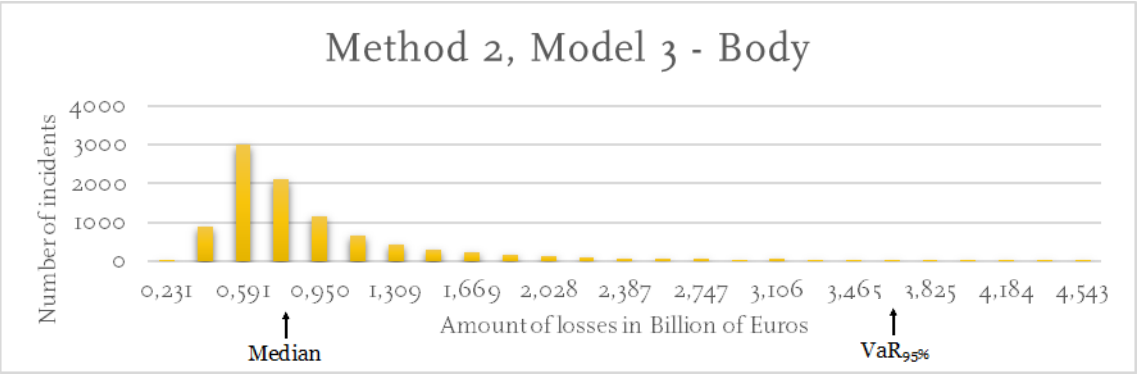

Method 1, Model 8 - Tail
Number of incidents
80
60
40
20
0
0,705
0,733
0,761
0,790
0,818
0,846
0,875
0,903
0,931
0,960
0,988
1,016
Amount of losses in Billion of Euros

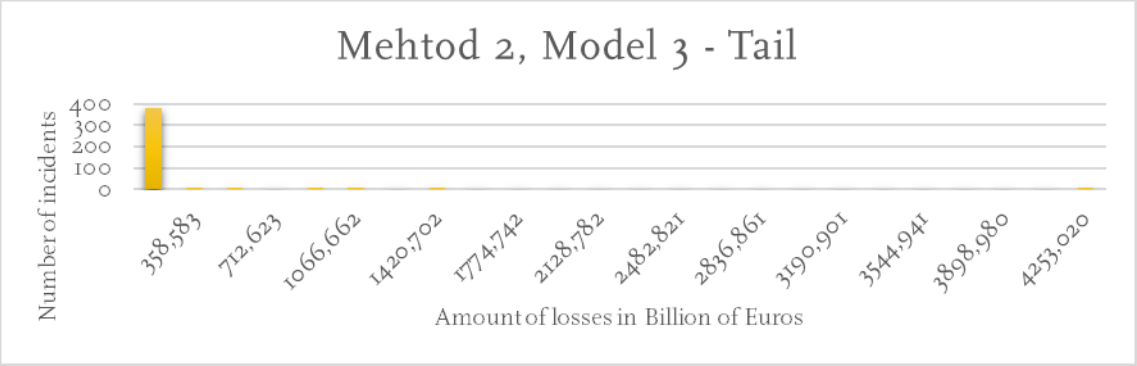

Method 2, Model 1 - Body
Number of incidents
3000
2000
1000
0
0,246
0,582
0,918
1,253
1,589
1,925
2,260
2,596
2,932
3,268
3,603
3,939
4,275
Amount of losses in Billion of Euros
Median
VaR95%

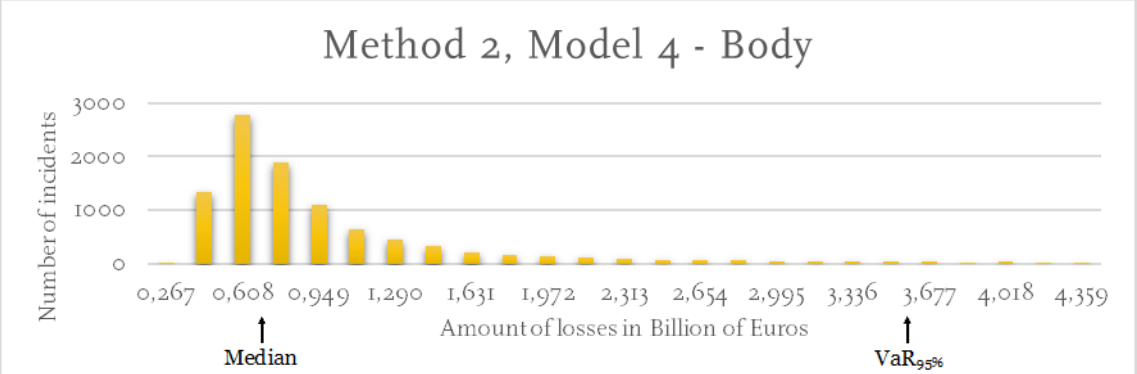

Method 2, Model 1 - Tail
Number of incidents
500
400
300
200
100
0
4,275
994,818
1985,362
2975,905
3966,449
4956,992
5947,536
6938,079
7928,623
8919,166
9909,710
10900,254
11890,797
Amount of losses in Billion of Euros

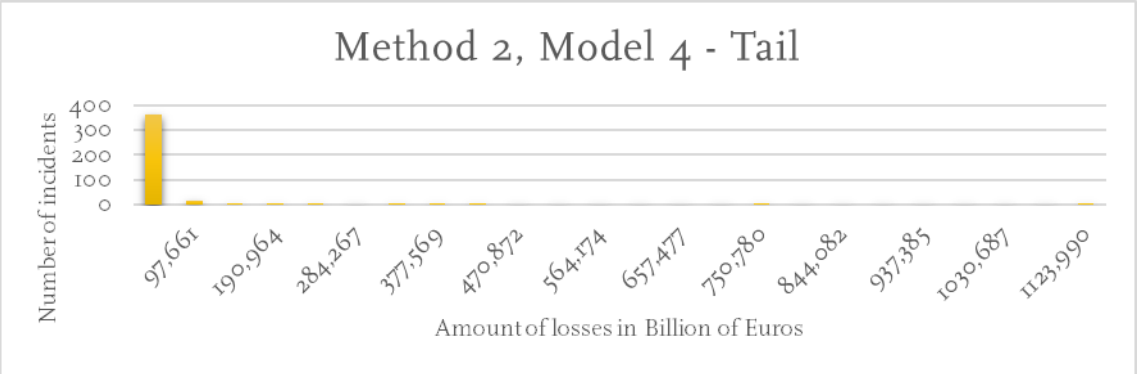

Method 2, Model 2 - Body
Number of incidents
3000
2000
1000
0
0,266
0,594
0,923
1,252
1,581
1,910
2,239
2,568
2,897
3,226
3,555
3,884
4,212
Amount of losses in Billion of Euros
Median
VaR95%

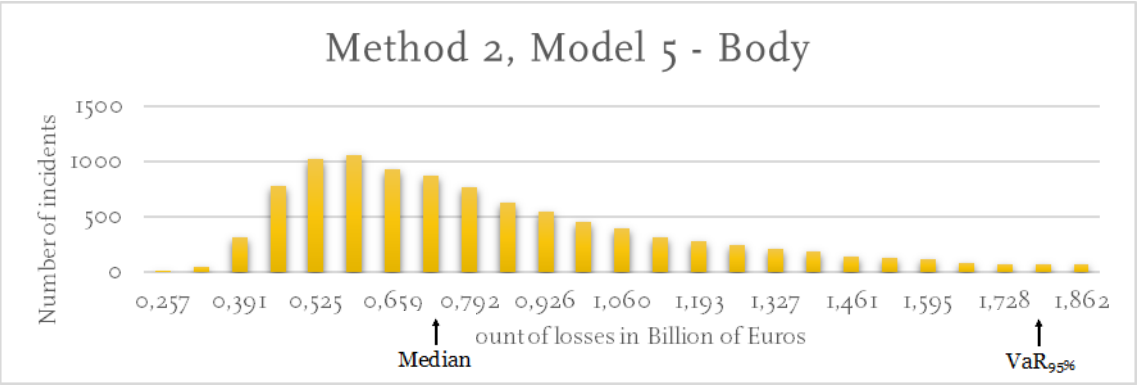

Method 2, Model 2 - Tail
Number of incidents
500
400
300
200
100
0
529,225
1054,237
1579,249
2104,262
2629,274
3154,286
3679,299
4204,311
4729,323
5254,335
5779,348
6304,360
Amount of losses in Billion of Euros

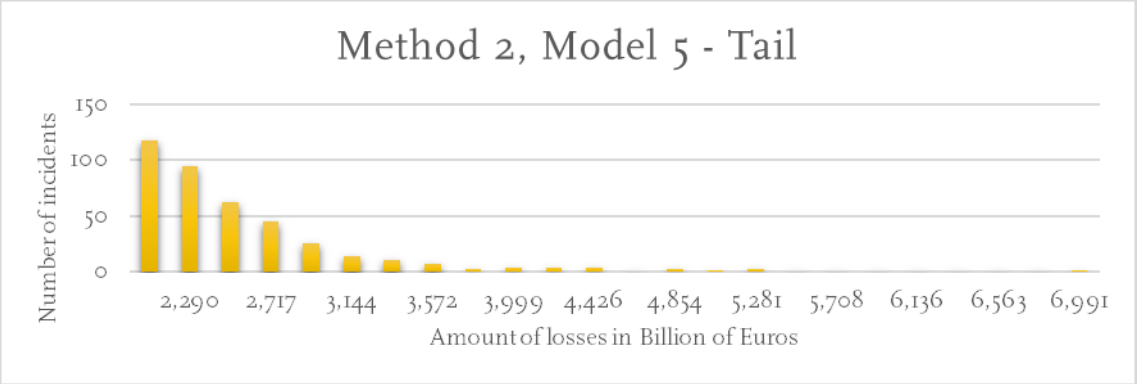

Method 2, Model 3 - Body
Number of incidents
4000
3000
2000
1000
0
0,231
0,591
0,950
1,309
1,669
2,028
2,387
2,747
3,106
3,465
3,825
4,184
4,543
Amount of losses in Billion of Euros
Median
VaR95%

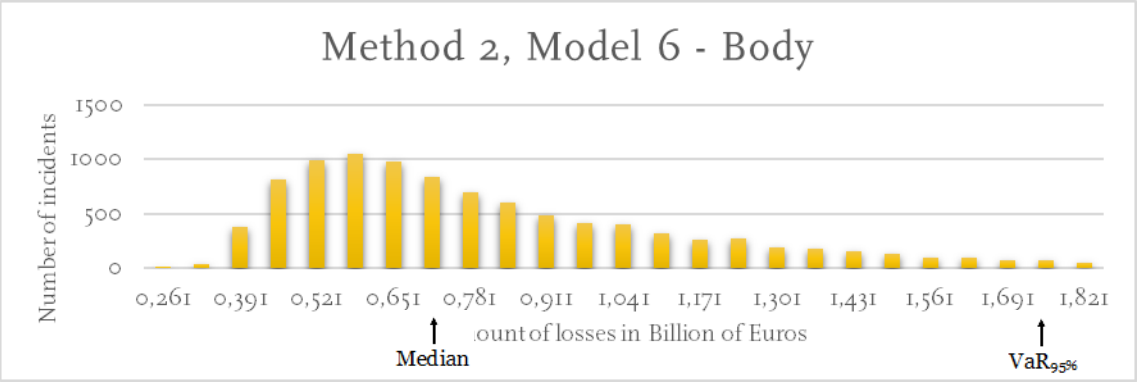

Mehtod 2, Model 3 - Tail
Number of incidents
400
300
200
100
0
358,583
712,623
1066,662
1420,702
1774,742
2128,782
2482,821
2836,861
3190,901
3544,941
3898,980
4253,020
Amount of losses in Billion of Euros

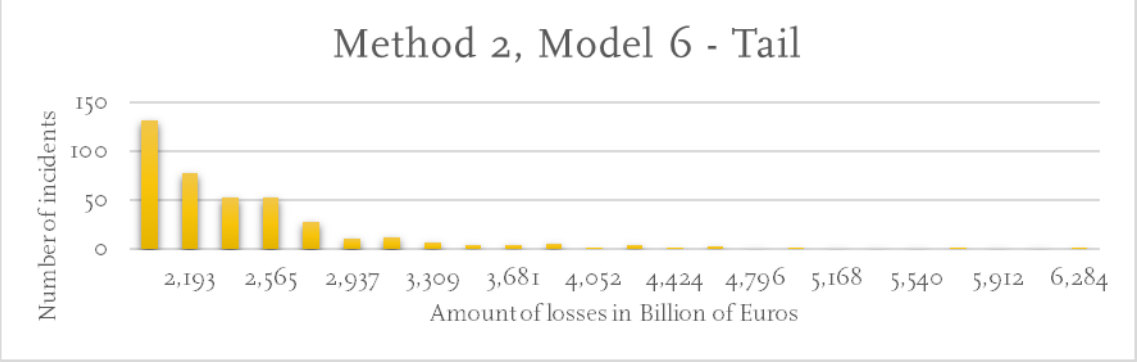

Method 2, Model 4 - Body
Number of incidents
3000
2000
1000
0
0,267
0,608
0,949
1,290
1,631
1,972
2,313
2,654
2,995
3,336
3,677
4,018
4,359
Amount of losses in Billion of Euros
Median
VaR95%

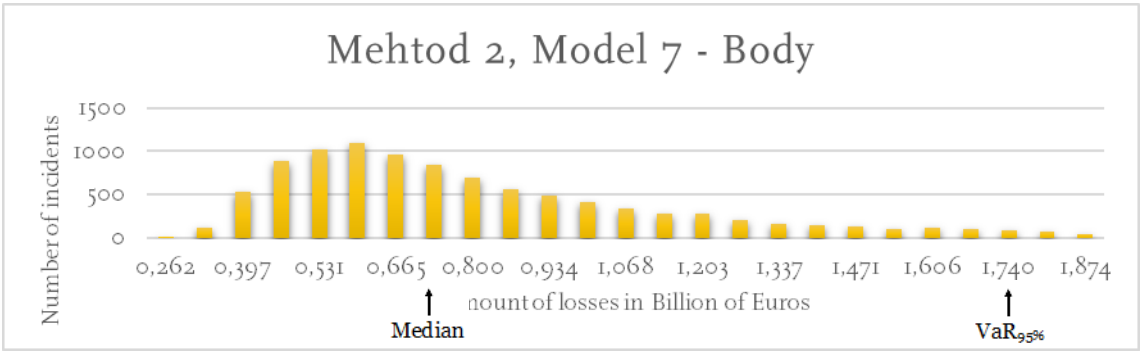

Method 2, Model 4 - Tail
Number of incidents
400
300
200
100
0
97,661
190,964
284,267
377,569
470,872
564,174
657,477
750,780
844,082
937,385
1030,687
1123,990
Amount of losses in Billion of Euros

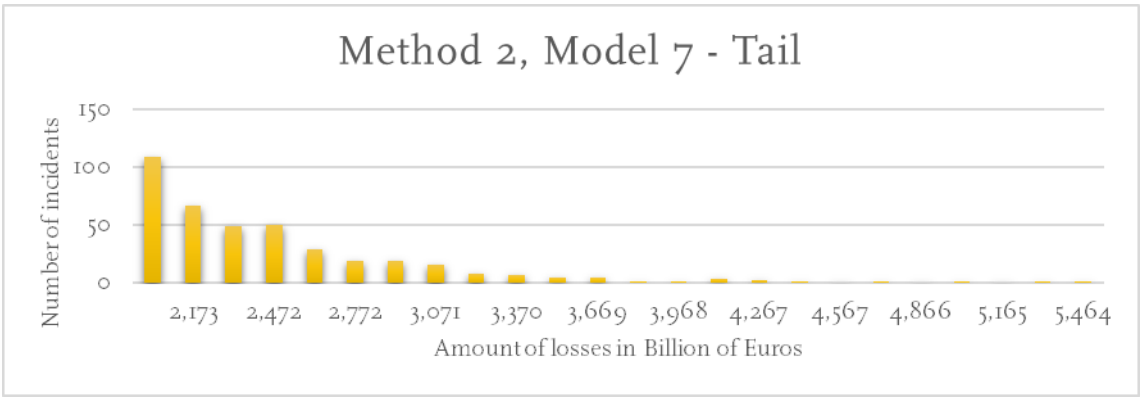

Method 2, Model 5 - Body
Number of incidents
1500
1000
500
0
0,257
0,391
0,525
0,659
0,792
0,926
1,060
1,193
1,327
1,461
1,595
1,728
1,862
ount of losses in Billion of Euros
Median
VaR95%

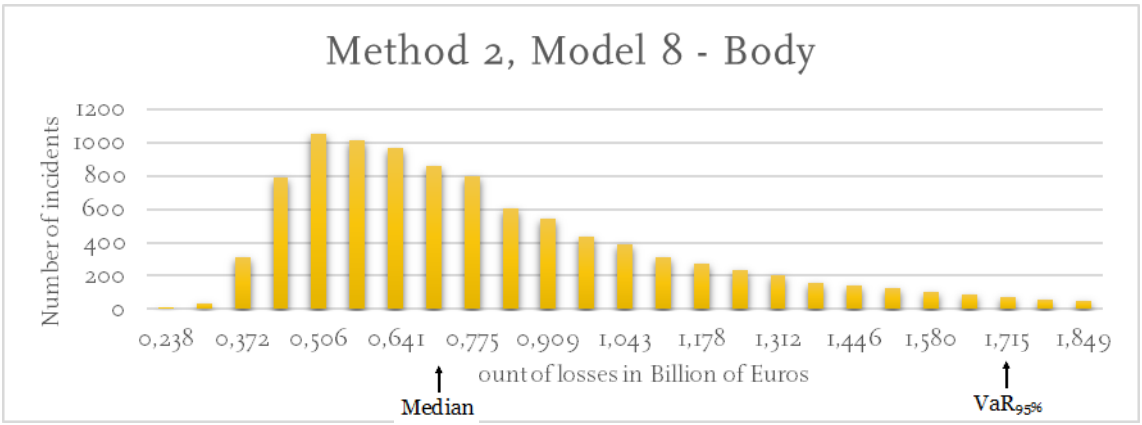

Method 2, Model 5 - Tail
Number of incidents
150
100
50
0
2,290
2,717
3,144
3,572
3,999
4,426
4,854
5,281
5,708
6,136
6,563
6,991
Amount of losses in Billion of Euros

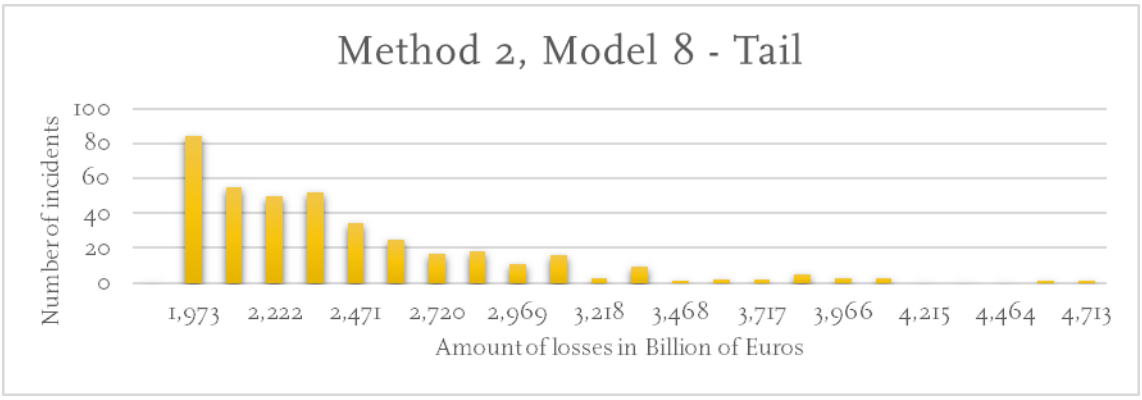

Method 2, Model 6 - Body
Number of incidents
1500
1000
500
0
0,261
0,391
0,521
0,651
0,781
0,911
1,041
1,171
1,301
1,431
1,561
1,691
1,821
ount of losses in Billion of Euros
Median
VaR95%

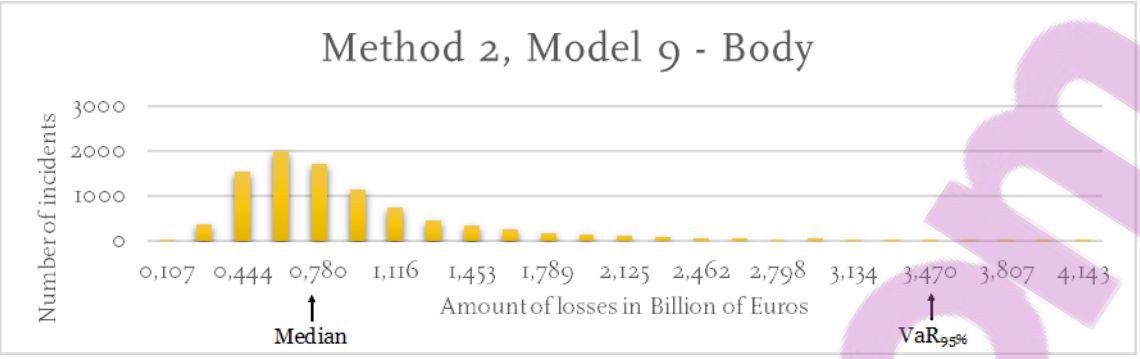

Method 2, Model 6 - Tail
Number of incidents
150
100
50
0
2,193
2,565
2,937
3,309
3,681
4,052
4,424
4,796
5,168
5,540
5,912
6,284
Amount of losses in Billion of Euros

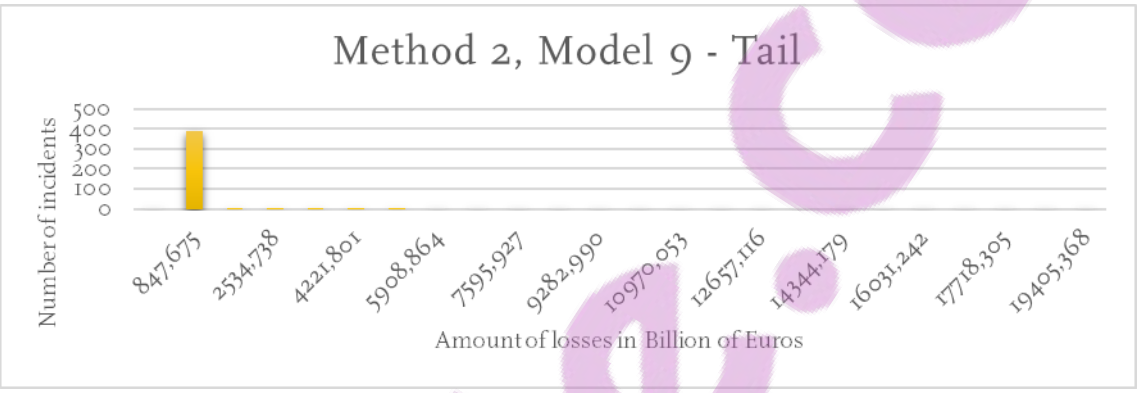

Mehtod 2, Model 7 - Body
Number of incidents
1500
1000
500
0
0,262 0,397 0,531 0,665 0,800 0,934 1,068 1,203 1,337 1,471 1,606 1,740 1,874
nount of losses in Billion of Euros
Median
VaR95%

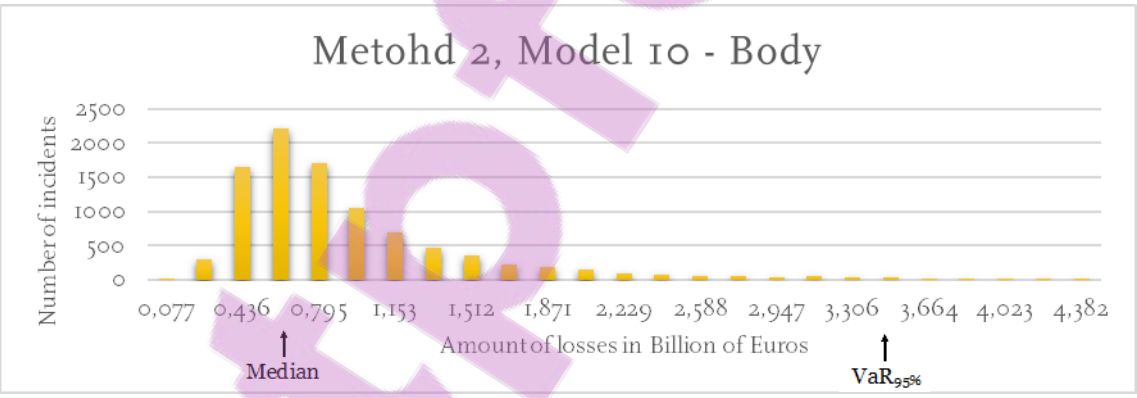

Method 2, Model 7 - Tail
Number of incidents
150
100
50
0
2,173 2,472 2,772 3,071 3,370 3,669 3,968 4,267 4,567 4,866 5,165 5,464
Amount of losses in Billion of Euros


Method 2, Model 8 - Body
Number of incidents
1200
1000
800
600
400
200
0
0,238 0,372 0,506 0,641 0,775 0,909 1,043 1,178 1,312 1,446 1,580 1,715 1,849
ount of losses in Billion of Euros
Median
VaR95%

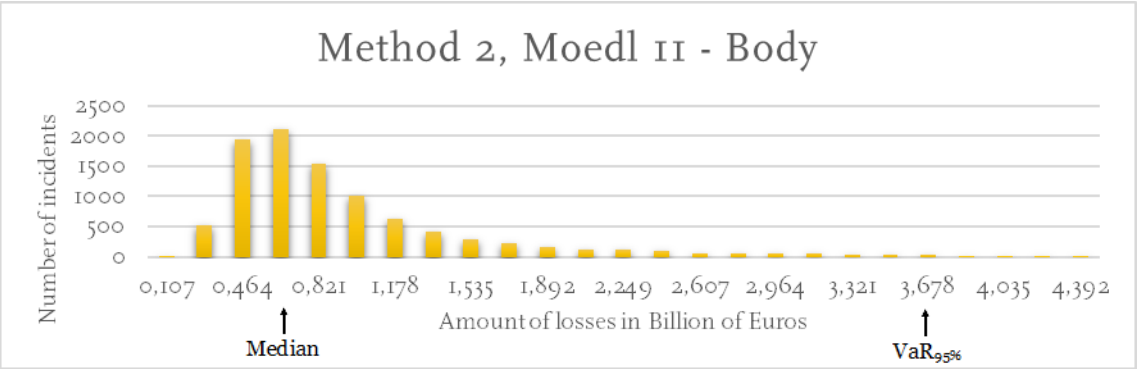

Method 2, Model 8 - Tail
Number of incidents
100
80
60
40
20
0
1,973 2,222 2,471 2,720 2,969 3,218 3,468 3,717 3,966 4,215 4,464 4,713
Amount of losses in Billion of Euros

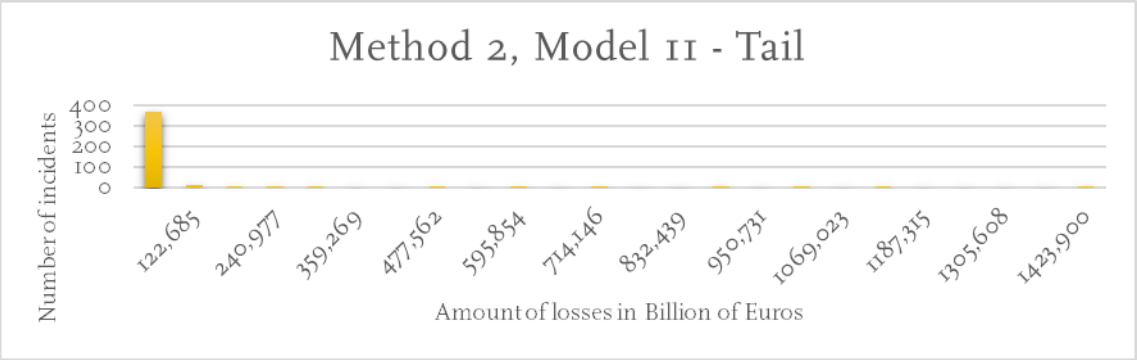

Method 2, Model 9 - Body
Number of incidents
3000
2000
1000
0
0,107
0,444
0,780
1,116
1,453
1,789
2,125
2,462
2,798
3,134
3,470
3,807
4,143
Amount of losses in Billion of Euros
Median
VaR95%

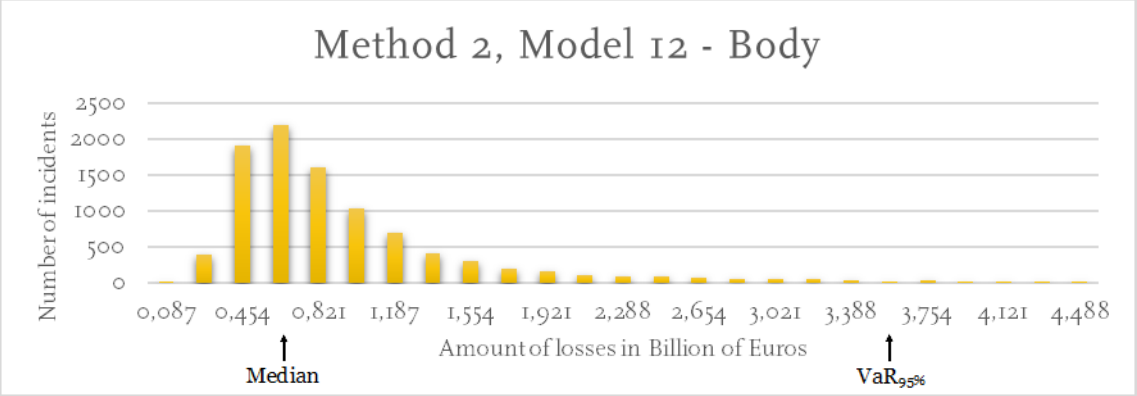

Method 2, Model 9 - Tail
Number of incidents
500
400
300
200
100
0
847,675
2534,738
4221,801
5908,864
7595,927
9282,990
10970,053
12657,116
14344,179
16031,242
17718,305
19405,368
Amount of losses in Billion of Euros

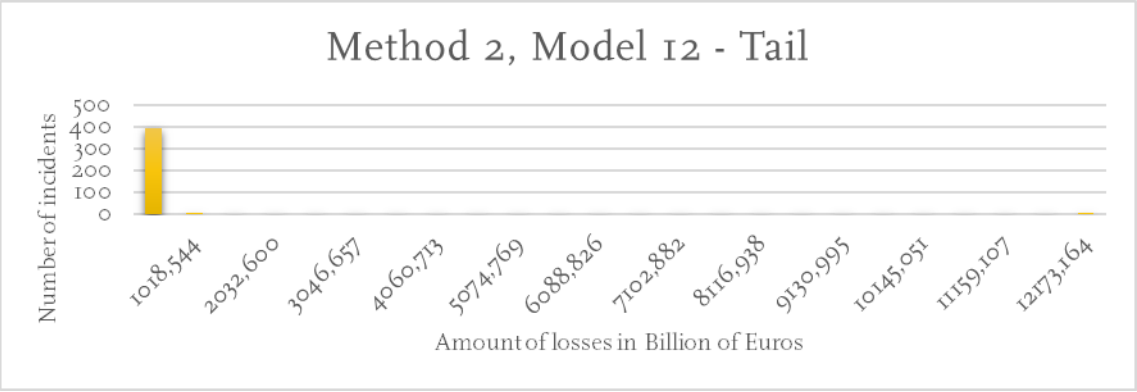

Metohd 2, Model 10 - Body
Number of incidents
2500
2000
1500
1000
500
0
0,077
0,436
0,795
1,153
1,512
1,871
2,229
2,588
2,947
3,306
3,664
4,023
4,382
Amount of losses in Billion of Euros
Median
VaR95%

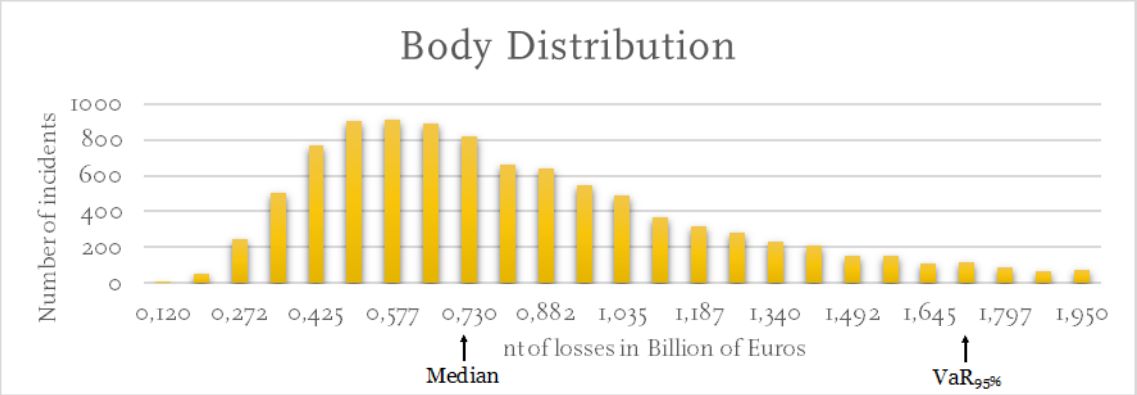

Method 2, Model 10 - Tail
Number of incidents
500
400
300
200
100
0
11444,391
22884,401
34324,411
45764,420
57204,430
68644,440
80084,449
91524,459
102964,468
114404,478
125844,488
137284,497
Amount of losses in Billion of Euros

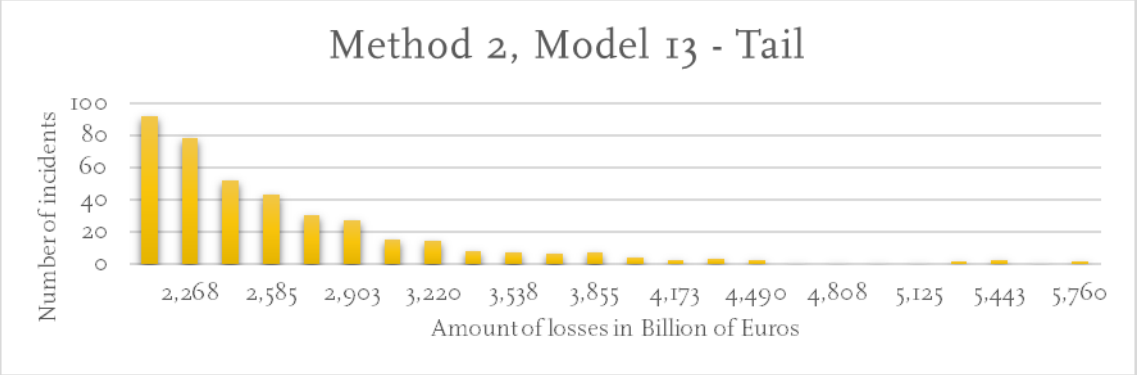

Method 2, Moedl 11 - Body
Number of incidents
2500
2000
1500
1000
500
0
0,107
0,464
0,821
1,178
1,535
1,892
2,249
2,607
2,964
3,321
3,678
4,035
4,392
Median
Amount of losses in Billion of Euros
VaR95%

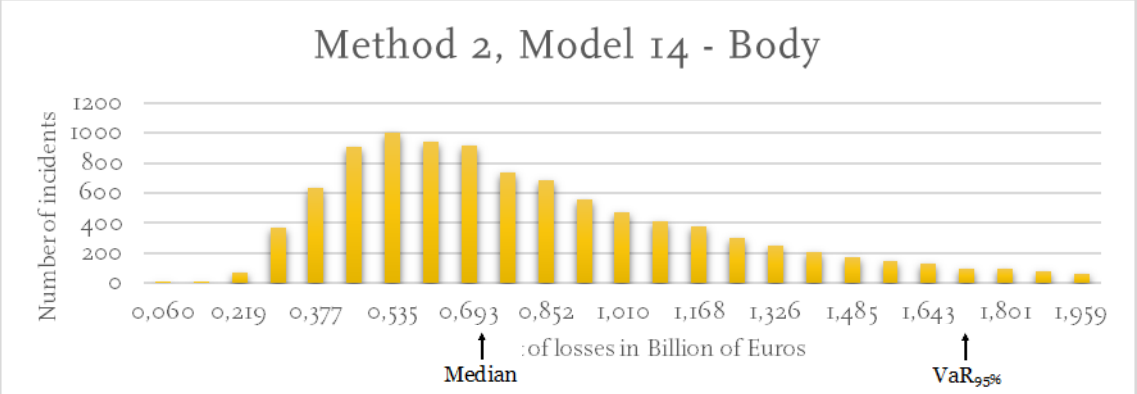

Method 2, Model 11 - Tail
Number of incidents
400
300
200
100
0
122,685
240,977
359,269
477,562
595,854
714,146
832,439
950,731
1069,023
1187,315
1305,608
1423,900
Amount of losses in Billion of Euros

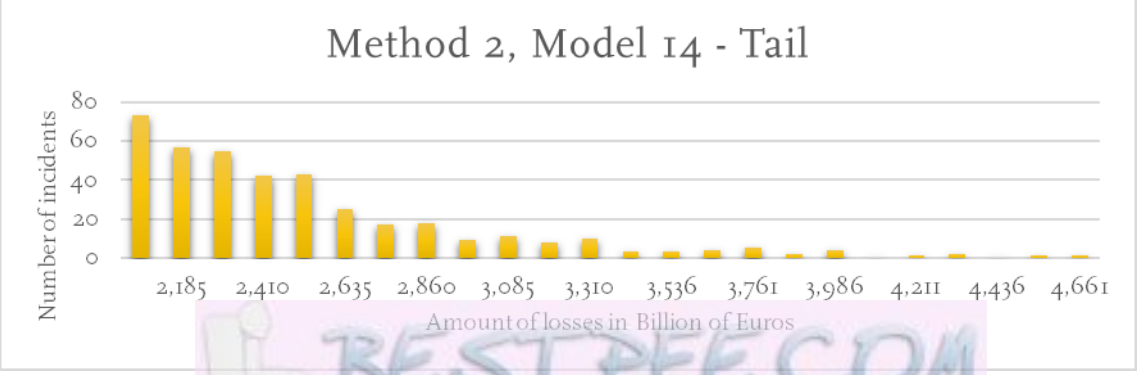

Method 2, Model 12 - Body
Number of incidents
2500
2000
1500
1000
500
0
0,087
0,454
0,821
1,187
1,554
1,921
2,288
2,654
3,021
3,388
3,754
4,121
4,488
Median
Amount of losses in Billion of Euros
VaR95%

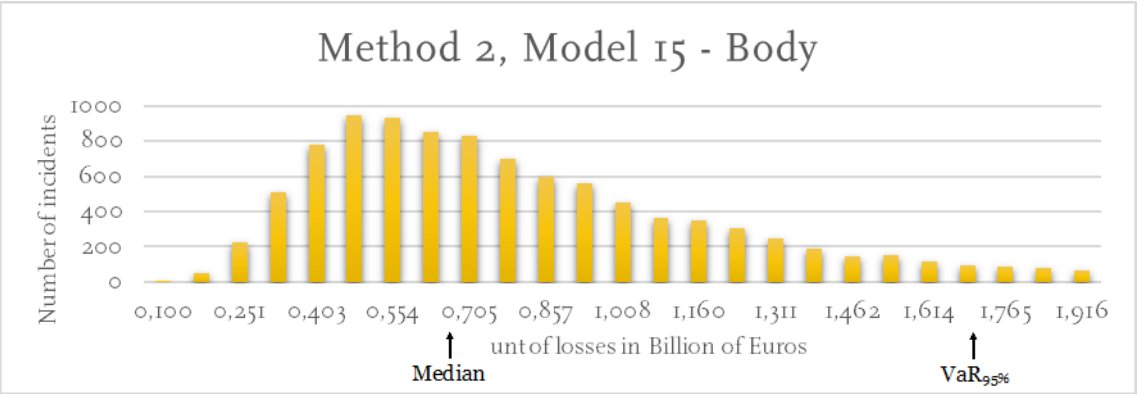

Method 2, Model 12 - Tail
Number of incidents
500
400
300
200
100
0
1018,544
2032,600
3046,657
4060,713
5074,769
6088,826
7102,882
8116,938
9130,995
10145,051
11159,107
12173,164
Amount of losses in Billion of Euros

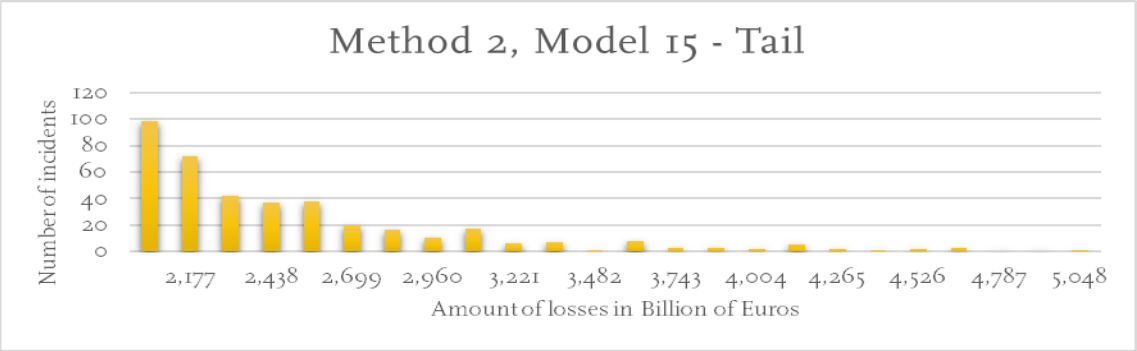


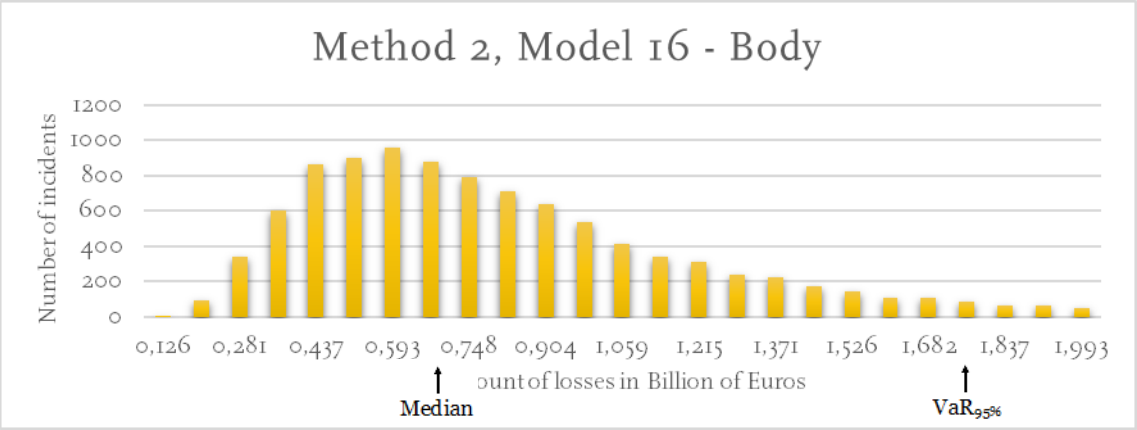



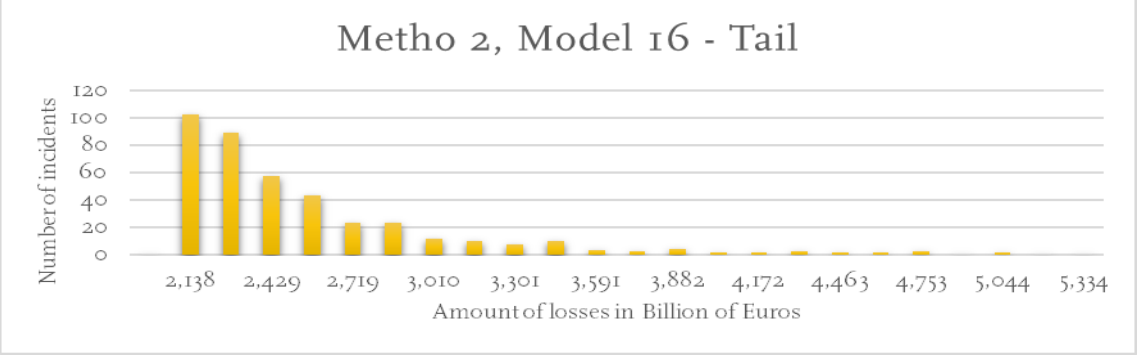




Method 2, Model 15 - Body
Number of incidents
1000
800
600
400
200
0
0,100 0,251 0,403 0,554 0,705 0,857 1,008 1,160 1,311 1,462 1,614 1,765 1,916
unt of losses in Billion of Euros
Median
VaR95%


Method 2, Model 15 - Tail
Number of incidents
120
100
80
60
40
20
0
2,177 2,438 2,699 2,960 3,221 3,482 3,743 4,004 4,265 4,526 4,787 5,048
Amount of losses in Billion of Euros


Method 2, Model 16 - Body
Number of incidents
1200
1000
800
600
400
200
0
0,126 0,281 0,437 0,593 0,748 0,904 1,059 1,215 1,371 1,526 1,682 1,837 1,993
ount of losses in Billion of Euros
Median
VaR95%


Metho 2, Model 16 - Tail
Number of incidents
120
100
80
60
40
20
0
2,138 2,429 2,719 3,010 3,301 3,591 3,882 4,172 4,463 4,753 5,044 5,334
Amount of losses in Billion of Euros

*Appendix 2 - Method Two's Models and their Distribution-assumptions*

Below is a table showing the models in method two and their exact composition of distribution assumptions.

| Model | Frequency distribution | Severity-body distribution | Severity-tail distribution |
|---|---|---|---|
| **1** | Poisson | Pearson 5 | Pareto |
| **2** | Poisson | Inverse Gaussian | Pareto |
| **3** | Poisson | Lognormal | Pareto |
| **4** | Poisson | Log-Logistic | Pareto |
| **5** | Negative Binominal | Pearson 5 | Pareto |
| **6** | Negative Binominal | Inverse Gaussian | Pareto |
| **7** | Negative Binominal | Lognormal | Pareto |
| **8** | Negative Binominal | Log-Logistic | Pareto |
| **9** | Poisson | Pearson 5 | Inverse Gaussian |
| **10** | Poisson | Inverse Gaussian | Inverse Gaussian |
| **11** | Poisson | Lognormal | Inverse Gaussian |
| **12** | Poisson | Log-Logistic | Inverse Gaussian |
| **13** | Negative Binominal | Pearson 5 | Inverse Gaussian |
| **14** | Negative Binominal | Inverse Gaussian | Inverse Gaussian |
| **15** | Negative Binominal | Lognormal | Inverse Gaussian |
| **16** | Negative Binominal | Log-Logistic | Inverse Gaussian |