CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Michal Spišiak

Assessment of cyber risk in the banking industry

Bachelor thesis

Prague 2017

Author: Michal Spišiak

Supervisor: doc. PhDr. Petr Teplý, Ph.D.

Academic Year: 2016/2017

Bibliographic note

SPIŠIAK, Michal. *Assessment of cyber risk in the banking industry.* Prague 2017. 56 pp. Bachelor thesis (Bc.) Charles University, Faculty of Social Sciences, Institute of Economic Studies. Thesis supervisor doc. PhDr. Petr Teplý, Ph.D.

Abstract

There has never been more need to discuss cybersecurity related issues. We live in a world where criminals do not have to physically visit a bank to steal money from it, where elections results can be influenced by data breached from personal email accounts, where to win a war a country needs skilled cybersecurity specialists rather than powerful weapons and where patients do not get recommended treatment because a hospital is under a cyberattack. The financial industry as a backbone of any modern economy requires adequate protection against cybercrime. We discuss major cyber threats for financial institutions as well as possible protection methods. After that we introduce Basel II Framework for operational risk assessment and we evaluate data breach risk in an empirical analysis.

Abstrakt

Nikdy nebylo důležitější diskutovat o věcech týkajících se kybernetické bezpečnosti než teď. Žijme ve světě, kde zločinci nemusí fyzicky navštívit banku, aby z ní mohli ukrást peníze, kde výsledky voleb můžou být ovlivněny uniknutími daty z osobních emailových účtů, kde na vyhrání války země potřebuje kvalifikované počítačové specialisty více než silné zbraně a kde pacienti nedostanou doporučené ošetření, protože nemocnice je pod kybernetickým útokem. Finanční odvětví jakožto páteř každé moderní ekonomiky vyžaduje přiměřenou ochranu proti kybernetickým trestním činům. Zabýváme se hlavními kybernetickými hrozbami pro finanční instituce a možnostmi

ochrany. Následně představíme Basel II Framework pro ohodnocení operačního rizika a vyhodnotíme riziko úniku dat v empirické analýze.

Keywords

operational risk, cyber risk, cybersecurity, data breach, bank

Klíčová slova

operační riziko, kybernetické riziko, kybernetická bezpečnost, únik dat, banka

Declaration of Authorship

1. The author hereby declares that he compiled this thesis independently, using only the listed resources and literature.
2. The author hereby declares that all the sources and literature used have been properly cited.
3. The author hereby declares that the thesis has not been used to obtain a different or the same degree.

Prague, 19 May 2017

Signature

Acknowledgment

I would like to express my gratitude to doc. PhDr. Petr Teplý, Ph.D. for his comments and supervision of my thesis.

Last but not least, I would like to thank to my parents who have been continuously supporting me in my study efforts by all means.

Contents

Introduction

One of the main concepts of cybersecurity is trust. This work explains how to develop this concept into reality.

This bachelor thesis is organized as follows. In the first part we discuss cyberattacks and their different forms. We explain most common ways in which attackers pursue their goals. One of the most valuable contributions is an easy to understand explanation of protection methods which can be used by banks and other organisations to protect themselves against cybercrime. Moreover, we provide examples of devastating cyberattacks which should motivate the reader to think about cyberattacks as about serious risks for the banking industry.

Next, we describe methodology we use. The model is inspired by typical applications of Basel II Framework. We explain how different frequency and severity distributions can be fitted to operational risk losses data, further we examine extreme value theory and its applications to the modelling of operational risk. We discuss assumptions of model as well.

Our empirical research is related to the risk of data breach which is one of the main components of cyber risk and it has the potential to destroy a bank. We propose hypotheses about frequency and severity distributions of data breaches. Then we use established and robust methodology which we previously described to test these hypotheses. Finally, we use risk measures to evaluate the the risk of data breach.

One of the largest contributions is that we apply established, robust and traditional methods to absolutely new dataset in the area of science where these methods have not been commonly used. Nonetheless, we pay enormous attention that the methods we use are suitable for solving the problem in front of which we stand. Cyber risk is for banks still relatively new topic and we are among the first researchers who write about it.

1 Cyberattacks

1.1 Threats

Cyberattacks are inevitable, however organisations[1] can mitigate their number and consequences by issuing proper security policies. The first step when thinking about protecting information systems of an organisation should be to identify threads. By this we mean identifying a class of unknown threats rather than particular people or institutions. These unknown classes of threats can be either individuals, other companies, governments or groups of them. Each class of threats has different objectives and requires different approach to protect against it.

For example some cybercriminals specialise in stealing money from ATMs, others involve in credit card frauds and still others demand money as a ransom for decrypting files. Another example are some companies which may want to weaken competition on the market and therefore they are not afraid of executing a cyberattack against their competitors. Yet another example are some governments which fight against terrorists and they use cyberattacks to get data from mobile devices used by these terrorists. Each of theses classes of threats is specific and protection against one of them might not work against others.

In the early days of the Internet, cybercriminals only wanted to gain respect of their community. Their attacks usually did not involve financial transactions, because Bitcoin did not exist and probably also because financial transactions in general were not very often executed over the Internet. Nonetheless, they were able to delete files on victim's computer and they could make the computer to run incredibly slowly too, which means they were already quite dangerous. When the Internet started to be widely used, their incentives have changed and they realised that cybercrime can earn some money to them. When the number of people connected to the Internet

[1] We often write about organisations when we analyse various cybersecurity issues or when we issue recommendations, because our arguments apply to a broader group of organisations than just banks. Nonetheless our main focus is on financial institutions.

increases then also the number of possible targets for cybercrime increases. When the ratio of connected people rises above some level, possible financial gain from a well planned attack can be higher than the amount of money invested into its preparation. This might be one of the main reasons why organised cybercriminal groups now exist.

1.2 Malware

Electronic Frontier Foundation (2017) explains: "Malware is short for malicious software: programs that are designed to conduct unwanted actions on your device. Computer viruses are malware. So are programs that steal passwords, secretly record you, or delete your data."

First virus called Morris worm appeared in 1988. (WeLiveSecurity 2017[d]) Any malware which spreads itself without a direct action from an attacker is called worm. Developers of operating systems and programmes installed on computers generally do not want their software to be compromised and they try to protect it. Attackers therefore have to find innovative ways how to overcome protection of operating system or programmes on the computer before they can install their malicious software.

Common ways how attackers pursue their goal are for example by finding vulnerabilities which they can exploit or by attracting users to install the malware themselves. The later method is sometimes called a scam because it uses a disability of some users to distinguish between real and fake content on the Internet. It can be done for example by creating a fake website which looks almost like the original, however this one is operated by the attacker. This website then offers to install seemingly useful programme which solves some problem for the user, however it is a malware. Scam is commonly used also for stealing login credentials for example to online banking.

Attackers and software vendors are playing a never ending game which goes as follows. The vendor publishes some software and attacker finds some vulnerability in it. Now the vulnerability is called zero-day vulnerability because the vendor does not know about it. Before the vendor finds out

about it, the attacker executes a cyberattack. With a little bit of effort and luck the vendor finds out about the vulnerability and fixes it. The time when no patch is available can be quite long, however eventually it is released or the software is abandoned altogether. The same or other attacker finds another vulnerability and the cycle closes.

1.3 Ransomware

Ransomware is malware which once installed on a device starts encrypting files saved in the storage devices connected to the device. Encrypted files cannot be opened and if victims want them to be decrypted, they have to pay some ransom. Attackers are often smart and they target attacks of this kind on organisations which are responsible for protecting data which if lost could cause extensive harm. Since ransomware can get on victim's device through a vulnerability for which does not yet exist a patch, the only almost bulletproof protection method is to keep a backup copy of important files on a storage device disconnected from the device most of the time.

There is huge moral dilemma related to ransomware. Some claim that a person or a company hit by a ransomware should pay if and only if they value their files more than is the cost of paying the ransom. Their adversaries say that if at least some people pay the ransom, they help to keep this malicious activity profitable for attackers. Therefore they suggest not to pay the ransom in any case. The situation gets even more complicated when we take into account that the attacker may not decrypt the files although the ransom is paid. In this case it would be necessary to evaluate the probability that files will not be decrypted even if ransom is paid based on experiences of other victims affected by the same or similar malware and take an action accordingly. Moreover, the game that attackers are playing is a little bit different. If they decrypt files once ransom is paid, they get trust, however they must spent additional resources to implement decryption feature of the malware. In the other case, they loose trust, however they save precious resources. In conclusion, it depends on the victims if they want to be egoistic

or altruistic and as a result they should make a decision based on this choice.

In March 2017 "WikiLeaks, the organization notorious for leaking highly secure government data, published a cache of documents that reportedly exposes tactics the CIA uses to hack into our devices. WikiLeaks released more than 8 700 documents and files." (CNET 2017[h]) On one hand it is great that WikiLeaks has published detailed data about vulnerabilities, because they could be finally resolved, on the other hand, it gave cybercriminals a powerful tool. It is true that WikiLeaks gave some time to manufacturers to resolve the vulnerabilities (CNET 2017[i]), nonetheless it could have waited with publishing of documents until these vulnerabilities were completely resolved.

Later, in April 2017 a vulnerability breached from National Security Agency (CNET 2017[b]) and in May 2017 ransomware combined with a worm based on this vulnerability attacked computers around the world including hospitals (WeLiveSecurity 2017[f]) and it affected "more than 200 000 devices in at least 150 countries." (CNET 2017[g])

1.4 Distributed denial-of-service attacks

Denial-of-service attacks are attacks with a purpose of making a computer infrastructure of a target unavailable. Distributed denial-of-service attacks or DDoS attacks are denial-of-service attacks for which multiple devices connected to the network are used. Typically some cybercriminals create on the fly a botnet which is a list of computers or devices connected to the Internet which they can control with a help of some malicious software. These are usually security cameras, consumer Wi-Fi routers and other Internet of Things devices without installed updates or with default passwords. (WeLiveSecurity 2017[a]) Another common example are web and other types of servers generally also without installed updates. When comes the right time or when somebody buys such botnet on a black market, all devices included in the botnet start floating servers of the target with requests. Any computer infrastructure which is providing some service has limited resources,

because practically all resources on Earth are limited, therefore if the amount of requests is too large, this infrastructure in the best case scenario stops responding to all request including legitimate and thus it becomes unavailable. Botnet can be also used to spread ransomware. (CNET 2017[a])

Commonly DDoS attacks are aimed against companies or governmental organisations which provide some service to customers or citizens. These types of attacks are usually not able to permanently damage the computer infrastructure of the target or cause a data breach, thus when the attack ends, system administrators can easily recover normal operation. Nonetheless, the length of the attack is usually in the hands of the attackers. The good news is that there usually exists relatively simple protection against DDoS attacks which requires increasing the number of simultaneous requests which the infrastructure is able to handle. The bad news is that this is usually very costly and usually only organisations operating critical information infrastructure are willing to accept this solution. Nevertheless, attackers most often do not receive any monetary benefit from executing a DDoS attack and on the contrary they risk legal punishment. Moreover, attackers can guilty for creating a botnet. (CNET 2017[e]) Therefore companies which are the most targeted are those with which the attackers have some ideological disputes.

2 Cyber risk

2.1 Legislation and regulation

If cybercriminals can stole 1 billion US dollars from a bank in one real example (Lab 2017[c]) and 951 million US dollars in another real example (Lab 2017[b]) then a cyber attack is clearly a risk for the bank.

Cyber risk is a part of operational risk. "Operational risk represents 5 – 30 % of banking risks, depending also on the extent to which it overlaps with the definition of other risks." (Mejstřík, Pečená, and Teplý 2015) Basel Committee on Banking Supervision (2006) in its Basel II Framework defined operational risk as "as the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events. This definition includes legal risk, but excludes strategic and reputational risk." Moreover, this document introduces three approaches for calculating operational risk capital requirement.

First, for risk capital K under the Basic Indicator Approach holds

$$K = \frac{\sum_{i=1}^{n} GI_i \cdot \alpha}{n},$$

where GI_i is equal to the annual gross income of the bank in year i if this income is positive and it is equal to zero, if this income is nonpositive, n is equal to three minus the number of past three years in which the annual gross income of the bank is nonpositive and α is equal to 15 %.

Second, under the Standardised Approach activities of the bank "are divided into eight business lines: corporate finance, trading & sales, retail banking, commercial banking, payment & settlement, agency services, asset management, and retail brokerage." (Basel Committee on Banking Supervision 2006) For risk capital K under this approach holds

$$K = \frac{\sum_{i=1}^{3} \max\left(\sum_{j=1}^{8} GI_i \cdot \beta_j, 0\right)}{3},$$

where GI_i is equal to the annual gross income of the bank in year i and β_j is for each j equal to a fixed percentage.

Third, under Advanced Measurement Approaches the bank can choose its own method to evaluate the operational risk. However, the selected method must be approved by a supervisor. (Basel Committee on Banking Supervision 2006) For example A. Chernobai, Jorion, and Yu (2011) use Advanced Measurement Approaches and adopt value at risk to calculate the risk capital.

Basel Committee on Banking Supervision (2016) later decided that the qualities of Advanced Measurement Approaches were exaggerated and started to prefer a standardised measure of operational risk. Such risk measure is Standardised Measurement Approach which uses mainly income and expenses of the bank for the calculation of its risk capital.

Government of the Czech Republic (2014) in its Act No. 181/2014 Coll. divides information system in the country into several categories based their significance for the country. Organisations operating information system from each category have different responsibilities such as providing contact information to an authority or detecting and alerting cyberattacks to the authority.

2.2 Corporate governance

There are protection methods which can be done only by the authorities. Still many protection methods can be implemented by banks and other companies themselves. Attackers often rely on social engineering when they design the cyberattack. Therefore it is crucial that employees are always cautious and they do not trust anybody on the Internet unless they verify the identity of the creator of the content witch which they are working.

This may sound too complicated, however to eliminated a majority of attacks it is enough that employees check whether their connection to a server is over SSL. Reality shows that even this may not be enough, because private key for the SSL certificate can be disclosed (WeLiveSecurity 2017[g]), however it is rather an exception than regularity. Another simple rule is to check if the URL of a website that is being visiting belongs to the organisation which

created the content on it. While the list of websites which are commonly visited by any employee does not change often, it should be easy to remember it. Content providers and system administrators can make this verification easier for users by issuing an extended validation SSL certificate for the website which allows to verify the owner of the website directly from an address bar of a browser.

Another simple advice is to instruct employees not to open email attachments from senders whose identity is unknown. Thanks to smart spam filters this task is considerably easier than in the past. There are two major ways how email attachments can damage the computer. First, this attachments is a malicious executable file which is not using any vulnerability, however there is an increased risk that the users accidentally allows its installation. Second, the attachment is a malicious executable file which is often hidden in an image or PDF file and which uses a vulnerability in common software installed on computers. (WeLiveSecurity 2017[h]) In this case it is not necessary to accept installation of this malware, nonetheless this type of malicious software is more difficult to create.

System administrators should periodically check if new updates are available and install them as soon as possible. Updates sometimes appear only a while before an attack, however they provide very strong protection against it. (CNET 2017[c]) To install a security software is likewise a good idea, nevertheless some security solutions based on machine learning can make things even worse. (WeLiveSecurity 2017[b])

Generally, running all applications with the lowest possible permissions can only increase security as well as enabling two-factor authentication if it is available. (CNET 2017[f]) Companies like Google understand that all win if attackers start working for them than working against them. Bounties for finding bugs are motivating security research as well. (WeLiveSecurity 2017[e])

2.3 Cybersecurity as a process

Cybersecurity is not a single security solution. It is a never-ending process which consists of infinite number of successes and the same number of failures. This is the inevitable reality. However nobody says that the ratio of successes to failures cannot be improved.

Suppose the first computer in the world including some software is created by a team of scientists and engineers. This team as a whole must understand each part of the computer and its software because otherwise they would not be able to produce it in the first place. It does not necessarily mean that there are no vulnerabilities in the computer and in its software. We can nonetheless say that if the team is cautious enough then it is capable of creating a system without vulnerabilities. Furthermore, by vulnerabilities we mean this time known vulnerabilities and also unknown vulnerabilities which are sometimes called zero-day vulnerabilities.

Now suppose that new computers with improved hardware and new pieces of software are developed. There is no need to reinvent the wheel and this principle can be perfectly applied in the computer science. Scientists and engineers who work on new hardware or software do not need to have the same knowledge as their predecessors. They need only a part of their predecessors' knowledge supplemented with some new information. As time goes on, as a result of this principle, computers and software are considerably diverse and each computer itself and each piece of software are composed from plenty of components. It is convenient that a creator of a new component does not need to have a perfect knowledge about all other components. Moreover, modern software development strategies prefer this approach because it lowers the requirement on human capital in terms of knowledge and therefore it speeds-up development and it also makes this process cheaper. The pace with which innovations come to the market today would not be possible if innovators had to know not everything about the inner working of components which they used in their product. It is enough that they know how to use these components.

This is on one hand a huge advantage, on the other hand it is a gigantic thread when it comes to cybersecurity. It does not hold anymore that the team of people which created this new innovative product understands every part of it. Therefore even if it was in the intention of the team to avoid any vulnerabilities they cannot avoid the risk of including some vulnerabilities in their product. They have to trust the producers of used components that they are secure.

Buyers of some device or software therefore have to trust entities about which they might not even know that they exist. If a vulnerability is known then there are two options. Either the producers of the affected component does not provide a patch for the vulnerability or it does provide it and the manufacturer of the final product does not deliver it to the consumer. Routers are among devices with plenty of vulnerabilities. (WeLiveSecurity 2017[i]) At first it might sound strange that the manufacturer of the final product is not interested in providing the patch to the consumer despite it is already available from the manufacturer of the affected component. However, there might be serious obstacles. For example, the final product is not designed to receive updates, because it would take additional time and resources to integrate such feature into the product. Similarly, if the product is not being sold anymore then the producers does not have an intention to spend additional time and resources to provide security updates. In a highly competitive market it is necessarily a disadvantage to spend these resources by a producers because they would definitely increase production cost. It does not matter who does not participate in the process of resolving the vulnerability, the consumer is always unprotected.

The worst case is if there is a zero-day vulnerability. Nonetheless, the producer cannot do anything in this case. Situation is strange when the government knows about these vulnerabilities, nonetheless it does not inform software producers about them. (CNET 2017[d]) The only possible solution is prevention. Manufacturers should sufficiently test their product for possible vulnerabilities before they start selling the product and they should not stop

researching vulnerabilities after the product is launched onto the market. For instance, if it is discovered that a similar product has a vulnerability, then this product is also suspicious to have similar vulnerability and it is thus reasonable for the manufacturer to follow up information about vulnerabilities in related products and take an appropriate action if it is necessary. This situation cannot be solved without government regulation or consumer awareness which would force producers to care more about security, which in turn is barely achievable without support from the government. Hopefully, more frequent cyberattacks will help to change something.

A nice example of a piece of software which is being developed for a long time by different teams of individuals and which consists of multiple interrelated components is operating system Linux. New operating system which puts security first and which is not based on Linux appeared recently. (Lab 2017[a]) Its creators have huge expectations, however only time will show whether new operating system, even if it is backed and supported by a strong company, can still get a reasonable market share or if the structure of the market is already determined for several centuries.

3 Methodology

The analysis of the cyber risk presented in this work makes use of an operational risk model which adheres to Advanced Measurement Approaches outlined by Basel II Framework. This analysis consists of finding both frequency and loss distributions of loss data and consequently estimating value at risk of an aggregate loss distribution which combines frequency and loss distributions into a single distribution. There are generally two approaches how to find a distribution which generated data that we work with. One option is to use a nonparametric approach and compute the empirical distribution function from definition. Another option would be to assume that the data follow a distribution from some family of distributions. By estimating unknown parameters we find one specific distribution. Furthermore, statistical tests can be used to infer whether the selection of distribution is reasonable or not. Although it is assumed that the reader is familiar with probability distributions used in this research, we provide an overview of them in order to avoid ambiguities with differences between commonly used parametrizations.

3.1 Frequency distributions

In operation risk modelling, frequency distributions are usually used to assign probabilities to different numbers of loss events per day. Discrete distributions are particularly useful for this task. We consider Poisson and negative binomial distributions.

Let X be a random variable with support $A \subset \mathbb{R}$. In the whole work we denote this by $X \in A$.

Poisson distribution It has a parameter $\lambda > 0$ and for any random variable X distributed according to the Poisson distribution we have

$$\begin{aligned} &X \in \{0, 1, 2, ... \}, \\ &P(X = j) = \frac{\lambda^j}{j!} e^{-\lambda}, \ j = 0, 1, 2, ... \\ &\mathrm{E}(X) = \lambda, \ \mathrm{Var}(X) = \lambda. \end{aligned}$$

If random variable X with Poisson distribution with parameter λ describes the number of loss events per one day, then random variable which describes the number of loss events per n days, has Poisson distribution with parameter $n\lambda$.

Negative binomial distribution It has two parameters $n \in \mathbb{N}$ and $p \in (0, 1)$ and for any random variable X distributed according to the negative binomial distribution we have

$$\begin{aligned} &X \in \{0, 1, 2, ... \}, \\ &P(X = j) = \binom{n + j - 1}{n - 1} p^n (1 - p)^j, \ j = 0, 1, 2, ... \\ &\mathrm{E}(X) = \frac{n(1-p)}{p}, \ \mathrm{Var}(X) = \frac{n(1-p)}{p^2}. \end{aligned}$$

The negative binomial distribution is a distribution of the number of failures before n-th success in a sequence of independent trials distributed according to Bernoulli distribution with parameter p.

3.2 Loss distributions

Before discussing particular distributions, let's review some properties of probability distributions which are relevant in the context of operational risk assessment.

Firstly, kurtosis is a property which is determined mostly by the tail of a probability distribution and for some random variable X it is usually calculated according to this formula:

$$\gamma_4 = \frac{\mu_4}{\sigma^4} = \frac{\mathrm{E}(X - \mathrm{E}(X))^4}{(\mathrm{E}(X - \mathrm{E}(X))^2)^2}.$$

Nonetheless, this is not the only way how kurtosis can be represented.

The reason why it is mentioned here is that explanation of kurtosis in many recent and older papers is misleading. For example, Balanda and MacGillivray (1988) define kurtosis as "the location- and scale-free movement of probability mass from the shoulders of a distribution into its center and tail," which is false and DeCarlo (1997) maintains that "many textbooks now correctly recognize that tailedness and peakedness are both components of kurtosis," what confirms that even more authors are claiming about kurtosis something what is not true. Westfall (2014) correctly asserts that kurtosis does not provide any relevant information about the peak of the distribution and provides examples of several distributions with the same mean, variance and kurtosis and shows that their peaks vary markedly.

Kurtosis of the normal distribution equals 3. Distribution which has larger kurtosis than the normal distribution is said to have positive excess kurtosis and it is called leptokurtic.

Similarly, by skewness we understand

$$\gamma_3 = \frac{\mu_3}{\sigma^3} = \frac{\mathrm{E}(X - \mathrm{E}(X))^3}{(\mathrm{E}(X - \mathrm{E}(X))^2)^{\frac{3}{2}}}.$$

Secondly, the tail of a distribution function $\mathrm{F}(x)$ is defined as $\overline{\mathrm{F}}(x) = 1 - \mathrm{F}(x)$. We say that a probability distribution with distribution function $\mathrm{F}(x)$ is heavy-tailed if the following condition holds for all $\alpha > 0$:

$$\lim_{x \to \infty} e^{\alpha x} \overline{\mathrm{F}}(x) = \infty.$$

By a loss distribution we mean a probability distribution of the number of breached records in a single loss event. The main purpose of the loss distribution is to describe the severity or intensity of losses. The severity of loss events (i.e. the number of breached records in our case) usually follows a distribution which is right skewed, has positive excess kurtosis, has heavy right tail and its support is on the positive values. (A. S. Chernobai, Rachev, and Fabozzi 2007) We fit the data to various distributions with these properties and for comparison also to some distributions which lack at least one of these properties. The right tail of the distribution is of particular

importance because this is the part of the distribution in which operational loss data differ the most from other kinds of data commonly used in many other fields. Modelling the severity of losses with distribution which is not heavy-tailed despite that these losses are actually following a heavy-tailed distribution can lead to consideraly underestimated operational risk.

Empirical distribution For a random sample $X_1, \dots, X_n$ the empirical distribution is defined with the following distribution function:

$$\hat{\mathrm{F}}_{\mathrm{n}}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\{(-\infty, x\rangle\}(X_i).$$

Its disadvantage is that it is not continuous.

Normal distribution It has two parameters $\mu \in \mathbb{R}$ and $\sigma > 0$ and it is not heavy-tailed. For any random variable X distributed according to the normal distribution we have

$$\begin{aligned} & X \in \mathbb{R}, \\ & \mathrm{E}(X) = \mu, \ \mathrm{Var}(X) = \sigma^2, \\ & \gamma_3 = 0, \ \gamma_4 = 3 \end{aligned}$$

and its density function is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}.$$

Exponential distribution It has a parameter $\lambda > 0$ and it is not heavy-tailed. For any random variable X distributed according to the exponential distribution we have

$$\begin{aligned} & X \in \langle 0, \infty), \\ & \mathrm{E}(X) = \frac{1}{\lambda}, \ \mathrm{Var}(X) = \frac{1}{\lambda^2}, \\ & \gamma_3 = 2, \ \gamma_4 = 9, \end{aligned}$$

its density function is

$$f(x) = \begin{cases} 0 & x < 0, \\ \lambda e^{-\lambda x} & x \geq 0 \end{cases}$$

and its distribution function is

$$f(x) = \begin{cases} 0 & x < 0, \\ 1 - e^{-\lambda x} & x \geq 0. \end{cases}$$

If random variable X describes the number of loss events per day and it has Poisson distribution with parameter λ, then the waiting time measured in days between two loss events has exponential distribution with parameter $\frac{1}{\lambda}$.

Gamma distribution It has a shape parameter $\alpha > 0$ and a scale parameter $\beta > 0$ and it is not heavy-tailed. For any random variable X distributed according to the gamma distribution we have

$$\begin{aligned} & X \in (0, \infty), \\ & \mathrm{E}(X) = \alpha\beta, \ \mathrm{Var}(X) = \alpha\beta^2, \\ & \gamma_3 = \frac{2}{\sqrt{\alpha}}, \ \gamma_4 = \frac{6}{\alpha} + 3 \end{aligned}$$

and its density function is

$$f(x) = \begin{cases} 0 & x \leq 0, \\ \frac{\beta^{-\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} & x > 0. \end{cases}$$

$\Gamma(y)$ is a function defined as $\Gamma(y) = \int_0^\infty t^{y-1} e^{-t}\, dt$ for $y \in (0, \infty)$.

Lognormal distribution It has two parameters $\mu \in \mathbb{R}$ and $\sigma > 0$ and it is heavy-tailed. For any random variable X distributed according to the lognormal distribution we have

$$\begin{aligned} & X \in (0, \infty), \\ & \mathrm{E}(X) = e^{\mu + \frac{\sigma^2}{2}}, \ \mathrm{Var}(X) = (e^{\sigma^2} - 1) e^{2\mu + \sigma^2}, \\ & \gamma_3 = \sqrt{e^{\sigma^2} - 1} \left(e^{\sigma^2} + 2 \right), \ \gamma_4 = 3e^{2\sigma^2} + 2e^{3\sigma^2} + e^{4\sigma^2} - 3 \end{aligned}$$

and its density function is

$$f(x) = \begin{cases} 0 & x \leq 0, \\ \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\log(x) - \mu)^2}{2\sigma^2}} & x > 0. \end{cases}$$

If Y is distributed according to the normal distribution, then $X = e^Y$ is distributed according to the lognormal distribution.

Weibull distribution It has a shape parameter $\alpha > 0$ and a scale parameter $\beta > 0$ and it is heavy-tailed only if $\alpha \in (0, 1)$. For any random variable X distributed according to the Weibull distribution we have

$$
\begin{aligned}
&X \in (0, \infty),\\
&\mathrm{E}(X) = \beta\Gamma\left(1 + \frac{1}{\alpha}\right),\ \mathrm{Var}(X) = \beta^2\left(\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma\left(1 + \frac{1}{\alpha}\right)^2\right),\\
&\gamma_3 = \frac{2\Gamma^3\left(1 + \frac{1}{\alpha}\right) - 3\Gamma\left(1 + \frac{2}{\alpha}\right)\Gamma\left(1 + \frac{1}{\alpha}\right) + \Gamma\left(1 + \frac{3}{\alpha}\right)}{\left(\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right)\right)^{\frac{3}{2}}},\\
&\gamma_4 = \frac{-3\Gamma^4\left(1 + \frac{1}{\alpha}\right) + 6\Gamma\left(1 + \frac{2}{\alpha}\right)\Gamma^2\left(1 + \frac{1}{\alpha}\right) - 4\Gamma\left(1 + \frac{3}{\alpha}\right)\Gamma\left(1 + \frac{1}{\alpha}\right) + \Gamma\left(1 + \frac{4}{\alpha}\right)}{\left(\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right)\right)^2},
\end{aligned}
$$

its density function is

$$
f(x) = \begin{cases} 0 & x \leq 0, \\ \alpha\beta^{-\alpha}x^{\alpha-1}e^{-\left(\frac{x}{\beta}\right)^{\alpha}} & x > 0 \end{cases}
$$

and its distribution function is

$$
f(x) = \begin{cases} 0 & x \leq 0, \\ 1 - e^{-\left(\frac{x}{\beta}\right)^{\alpha}} & x > 0. \end{cases}
$$

Pareto distribution It has a minimum value parameter $k > 0$ and a shape parameter $\alpha > 0$ and it is heavy-tailed. For any random variable X distributed according to the Pareto distribution we have

$$
\begin{aligned}
&X \in \langle k, \infty),\\
&\mathrm{E}(X) = \frac{\alpha k}{\alpha - 1} \text{ for } \alpha > 1 \text{ and does not exist otherwise,}\\
&\mathrm{Var}(X) = \frac{\alpha k^2}{(\alpha - 2)(\alpha - 1)^2} \text{ for } \alpha > 2 \text{ and does not exist otherwise,}\\
&\gamma_3 = \frac{2(\alpha + 1)}{\alpha - 3}\sqrt{\frac{\alpha - 2}{\alpha}} \text{ for } \alpha > 3 \text{ and does not exist otherwise,}\\
&\gamma_4 = \frac{3(\alpha - 2)\left(3\alpha^2 + \alpha + 2\right)}{(\alpha - 4)(\alpha - 3)\alpha} \text{ for } \alpha > 4 \text{ and does not exist otherwise,}
\end{aligned}
$$

its density function is

$$
f(x) = \begin{cases} 0 & x < k, \\ \alpha k^{\alpha}x^{-\alpha-1} & x \geq k \end{cases}
$$

and its distribution function is

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

Please notice that the support of a random variable with the Pareto distribution depends on the parameter k.

Cauchy distribution It has a location parameter $a \in \mathbb{R}$ and a scale parameter $b > 0$ and it is heavy-tailed. For any random variable X distributed according to the Pareto distribution we have

$$X \in \mathbb{R},$$

its density function is

$$f(x) = \frac{1}{b\pi}\left(1 + \left(\frac{(x-a)}{b}\right)^2\right)^{-1}$$

and its distribution function is

$$f(x) = \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{x-a}{b}\right).$$

Any moment of X does not exist.

3.3 Parameters estimation

Maximum likelihood estimation (MLE) is used in this research always when there is a need for parameter estimation. A short description of this method follows. Assume that a random sample $X = (X_1, \ldots, X_n)$ is from a distribution with a density function $f(x|\theta_X)$ such that $f(x|\theta_X) \in \{f(x|\theta) \colon \theta \in \Theta \subset \mathbb{R}^d\}$ where $d \in \mathbb{N}$ and θ is the parameter which should be estimated. The likelihood function is defined as $L_n(\theta) = \prod_{i=1}^{n} f(X_i|\theta)$. The maximum likelihood estimator $\hat{\theta}$ of the parameter θ_X is such a vector from Θ which maximises the likelihood function. In practice the log-likelihood function defined as $l_n(\theta) = \log(L_n(\theta)) = \sum_{i=1}^{n} log(f(X_i|\theta))$ is maximised because of computational simplicity. Maxima of likelihood and log-likelihood functions are in the same point as logarithm is an increasing function on its whole

domain. If several regularity conditions are met then the maximum likelihood estimators are consistent and asymptotically normally distributed. (Kulich 2014) This reason together with the availability of numerical methods for computing maximum likelihood estimates makes MLE very attractive to use.

3.4 Goodness of fit

Q-Q plot Suppose $\tau \in \langle 0, 1\rangle$. q_τ is τ-th quantile of a distribution of a random variable X with continuous distribution function if $P(X \leq q_\tau) = \tau$. Q-Q plot can be used to compare empirical quantiles with quantiles of a previously selected distribution. On the vertical axis are values of the selected distribution and on the horizontal axis are values of the empirical distribution. Q-Q plot then shows the relationship between τ-th quantiles of the empirical distribution and τ-th quantiles of the selected distribution. If τ-th quantiles of both distributions are the same for all $\tau \in \langle 0, 1\rangle$ then Q-Q plot is equivalent to the axis of the first quadrant.

An advantage of the Q-Q plot is that we can easily compare how these two distributions deviate in their tails with how much they deviate in their bodies. If the selected distribution underestimates empirical quantiles q_τ for τ from some interval then the part of Q-Q plot which displays quantiles q_τ for τ from this interval is above the axis of the first quadrant. A disadvantage of the Q-Q plot is that it does not provide any formal way to test if the differences between quantiles of these two distributions are significant.

Kolmogorov-Smirnov test The null hypothesis of this test is

$$\mathrm{F_X}(x) = \mathrm{F}(x) \text{ for all } x \in \mathbb{R},$$

where $\mathrm{F_X}$ is the distribution function of a possibly unknown distribution which generated a random sample $X_1, \ldots, X_1$ and $\mathrm{F}(x)$ is the distribution function of a distribution which we think is the distribution which generated this random sample.

The test statistic is

$$\mathrm{KS} = \sup_{x \in \mathbb{R}} \left| \hat{\mathrm{F}}_\mathrm{n}(x) - \mathrm{F}(x) \right|.$$

Rachev, Kim, and Bianchi (2011) provide a computing formula for the test statistic:

$$\mathrm{KS} = \max\left(\max_{1\leq i\leq n}\left(\frac{i}{n} - \mathrm{F}\left(X_{(i)}\right)\right), \max_{1\leq i\leq n}\left(\mathrm{F}\left(X_{(i)}\right) - \frac{i-1}{n}\right)\right),$$

where $X_{(k)}$ is the k-th smallest value among $X_1, \ldots, X_n$.

Anderson-Darling test The null hypothesis is the same as in the Kolmogorov-Smirnov test, however "unlike the Kolmogorov-Smirnov test, the Anderson-Darling test tends to put most weight on the tails of the distribution." (A. S. Chernobai, Rachev, and Fabozzi 2007)

The test statistic is

$$\mathrm{AD} = \int_{-\infty}^{\infty} \frac{\left(\widehat{\mathrm{F}}_{\mathrm{n}}(x) - \mathrm{F}(x)\right)^2}{\mathrm{F}(x)\left(1 - \mathrm{F}(x)\right)}\, d\,\mathrm{F}(x).$$

Original authors of the test Anderson and Darling (1954) provide also a computing formula for the test statistic:

$$\mathrm{AD} = -1 - \frac{1}{n^2}\sum_{i=1}^{n}(2i-1)\left(\log\left(\mathrm{F}\left(X_{(i)}\right)\right) - \log\left(1 - \mathrm{F}\left(X_{(n-i+1)}\right)\right)\right).$$

4 Extreme value theory

The central limit theorem is one of the most fundamental theorems of modern statistical theory and it finds its application in many fields of statistical research. However, there are some phenomena in the real world which cannot be described by the normal distribution, what means that the central limit theorem can be applied only to substantially smaller set of problems than we would wish. Many current world phenomena follow heavy-tailed distributions. Nair, Wierman, and Zwart (2013) suggest that so-called "catastrophe principle" of heavy-tailed distributions is so convenient that this class of distributions becomes increasingly more used by the research community. This principle states that if some random variables $X_1, \ldots, X_n$ follow a heavy-tailed distribution then for some number x we have $P(\max(X_1, \ldots, X_n) > x) \sim P(X_1 + \cdots + X_n > x)$. In principle it means that if random variables $X_1, \ldots, X_n$ represent losses then in any

realisation of an experiment, one of these random variables achieves so huge value that values of all other random variable are negligible. Especially, this property is very common for operational losses. Many losses recorded on daily basis are really insignificant, nonetheless when waiting sufficiently long a huge loss event appears and the stability of the bank is instantly threatened.

Extreme value theory can be used to model the most severe losses. Kemp (2011) indicates: "tail behaviour can only in practice take a small number of possible forms" and extreme value theory explains what these forms can be. It is an uneasy task to find the distribution of operational losses severity, nevertheless if the task is modified to finding only the tail of the distribution of operational losses severity then according to the previous claim of Kemp (2011) there is a chance that modelling can become simpler. Moreover, when modelling the severity of operational losses, tiny everyday losses are not in the center of our attention. We are concerned about huge losses in the right tail of the probability distribution which can cause serious troubles for the impacted institution and in the worst case scenario they can spread the contagnion. Disastrous loss may happen only once in a century, however if it happens the aftermaths for financial sector and consequently for the whole economy are enormous. Nevertheless, we should not be overly optimistic. This theory brings new possibilities how to accurately measure operational risk, however it is at a cost of additional assumptions. Statistical analysis is always like this: if you want to gain something, you have to give up something. Extreme value theory can be applied also in many other fields dealing with rare events, it comes handy for instance in studying of floods. (Einmahl, Li, and Liu 2009)

We are using point over threshold method to apply extreme value theory to dataset consisting of numbers of records breached in individual loss events. This method is working with losses which lie above a high threshold u. Therefore results depend on the selection of the threshold. Suppose X is a random variable whose support does not have an upper bound and u is a

large threshold. Then function

$$\mathrm{F}_u(x) = P(X - u \leq x | x > u) \text{ for } x \in \langle 0, \infty)$$

is called the excess distribution function of X over the threshold u.

Generalized Pareto distribution with a shape parameter $\xi \in \mathbb{R}$ and a scale parameter $\beta > 0$ is defined with the following tail of distribution function:

$$\overline{\mathrm{G}}_{\xi,\beta}(x) = \begin{cases} \left(1 + \xi \frac{x}{\beta}\right)^{-\frac{1}{\xi}} & \xi \neq 0, \\ e^{-\frac{x}{\beta}} & \xi = 0 \end{cases}$$

for $x \in \langle 0, \infty)$ if $\xi \geq 0$ and $x \in \langle 0, -\frac{\beta}{\xi})$ if $\xi < 0$. (Paul Embrechts, Klüppelberg, and Mikosch 1997) Parameter ξ indicates the weight of the tail of the distribution and parameter β represents volatility. (Bali 2003)

If X is a random variable whose support does not have an upper bound and u is a large threshold then the tail of the excess distribution function converges to the generalized Pareto distribution. Following the formal notation of Paul Embrechts, Klüppelberg, and Mikosch (1997) we have

$$\lim_{u \to \infty} \sup_{0 < x} |\overline{\mathrm{F}}_u(x) - \overline{\mathrm{G}}_{\xi,\beta(u)}(x)| = 0$$

or

$$\overline{\mathrm{F}}_u(x) \approx \overline{\mathrm{G}}_{\xi,\beta(u)}(x).$$

It is beneficial to note that the parameter $\beta(u)$ depends on the selection of the threshold u.

Suppose $X_1, \ldots, X_n$ is a random sample, u is a high threshold and excesses over the threshold defined as $X_i - u, i = 1, \ldots, n$ follow a generalized Pareto distribution. If there are N_u observations above the threshold u and ξ and β are estimates of the respective parameters then the estimator of the value at risk with confidence level $(1-\alpha)$ is provided by Paul Embrechts, Klüppelberg, and Mikosch (1997):

$$\widehat{\mathrm{VaR}}_{1-\alpha} = u + \frac{\hat{\beta}}{\hat{\xi}} \left(\left(\frac{n}{N_u} (1-\alpha)^{-\hat{\xi}} - 1 \right) \right)$$

and the estimator of the conditional value at risk (CVaR) with confidence

level $(1-\alpha)$ is provided by Chavez-Demoulin and P. Embrechts (2004):

$$\widehat{\mathrm{CVaR}}_{1-\alpha} = \left(\frac{1}{1-\hat{\xi}} + \frac{\hat{\beta} - \hat{\xi}u}{\left(1-\hat{\xi}\right)\widehat{\mathrm{VaR}}_{1-\alpha}} \right) \widehat{\mathrm{VaR}}_{1-\alpha}.$$

Unfortunately, we are not using these two formulas in the further text because they cannot be used to calculate value at risk of the aggregate loss distribution. Nonetheless, they explain how powerful the extreme value theory actually is. By just knowing the tail of the distribution function without potentially any knowledge about the rest of the distribution we are able calculate the value at risk of the distribution what is a property of the whole distribution. This is definitely not trivial.

The selection of the threshold u is critical because through the parameter $\beta(u)$ which depends on u, the tail of the distribution used for value at risk calculation is altered. Suppose X is a random variable whose support does not have an upper bound and $X_1, \ldots, X_n$ is a random sample. Mean excess function of X defined as

$$\mathrm{e}(u) = \mathrm{E}(X - u | X > u) \text{ for } u \in \langle 0, \infty)$$

and empirical mean excess function defined as

$$\hat{\mathrm{e}}_{\mathrm{n}}(u) = \frac{\sum_{i=1}^{n} (X_i - u)\, \mathbb{I}\{(u,\infty)\}(X_i)}{\sum_{i=1}^{n} \mathbb{I}\{(u,\infty)\}(X_i)} \text{ for } u \in \langle 0, \infty)$$

(Ghosha and Resnickb 2010) can be used to graphically decide which threshold to choose. Paul Embrechts, Klüppelberg, and Mikosch (1997) suggest to choose such threshold u that the empirical mean excess function $\hat{\mathrm{e}}_{\mathrm{n}}(x)$ is approximatelly linear for $x \geq u$. For example, if the empirical mean excess function is approximatelly linear with just few points in which it immensely changes its slope then the last point where it changes its slope is usually a decent choice of the threshold and above this threshold data approximatelly follow generalized Pareto distribution. There is no exact rule according to which a correct threshold can be chosen. We must accept the fact that any choice of the threshold can introduce inaccuracy in the final result. Even worse is that we will never find out how large this inaccuracy is. This is one of the largest disadvantages of the extreme value theory.

5 Risk measures

First we introduce how different risk measures can be compared qualitatively. Artzner et al. (1999) introduce a concept of coherence which acomplishes this task. Assume that Ω is a set containing all states of the nature and this set is finite. The set of all possible positive random variables defined on an appropriate probability space with a set of possible outcomes Ω is denoted by G. Please note that for any $X_1, X_2 \in G$ we have $X_1 + X_2 \in G$ and for any $X, \in G$, $a, b \in \mathbb{R}$ we have $aX + b \in G$, Risk measure ρ is a function which maps the set G to $\mathbb{R}$.

Risk measure ρ is called coherent if it satisfies the following four assumptions:

- (Translation invariance). For any $X \in G$ and any $a \in \mathbb{R}$ we have $\rho(X + a) = \rho(X) - a$;
- (Subadditivity). For any $X_1, X_2 \in G$ we have $\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$;
- (Positive homogeneity). For any $X \in G$ and any $a \in \langle 0, \infty)$ we have $\rho(aX) = a\rho(X)$;
- (Monotonicity). For any $X_1, X_2 \in G$ such that $X_1 \leq X_2$ we have $\rho(X_1) \leq \rho(X_2)$.

A special case of Translation invariance is $\rho(X + \rho(X)) = 0$. It means that adding sure amount $\rho(X)$ to a risky outcome X reduces the risk of the whole portfolio to zero. As a consequence, if a bank holds regulatory capital calculated according to a risk measure satisfying Translation invariance assumption then this regulatory capital offsets the risky outcome. Subadditivity of the risk measure guarantees that diversification reduces the risk. Positive homogeneity implies that doubling the size of the risky outcome X also doubles the risk.

One of the most used risk measures is value at risk (VaR). Suppose X is a random variable with a distribution function F and $\alpha \in \langle 0, 1 \rangle$. Value at

risk at confidence level $(1-\alpha)$ is

$$\mathrm{VaR}_{1-\alpha} = \inf\{x \in \mathbb{R} : \mathrm{F}(x) \geq 1-\alpha\} = \mathrm{F}^{-1}(1-\alpha)$$

where F^{-1} is a quantile function of the random variable X.

Value at risk can be intuitively interpreted as follows. Suppose X is a random variable representing a loss amount and we make n realisations of X. If n approaches infinity then $100 \cdot (1-\alpha)$ % of realisations of the loss amount do not exceed the value at risk at confidence level $(1-\alpha)$. Given that α is sufficientl small, value at risk measures the loss amount which is exceeded only in few cases.

Value at risk does not satisfy the Subadditivity assumption (Sarykalin, Serraino, and Stan Uryasev 2008). For this reason it is strictly criticised by Artzner et al. (1999), "value at risk does not behave nicely with respect to the addition of risks, even independent ones, thereby creating severe aggregation problems" and "the use of value at risk does not encourage and, indeed, sometimes prohibits diversification because value at risk does not take into account the economic consequences of the events, the probabilities of which it controls."

Another disadvantage of VaR mentioned by Lebovič (2012) and Angelidis and Degiannakis (2009) is that the value at risk does not provide any information about the size of the loss once it exceeds the $(1-\alpha)$-th quantile. If we remind that the largest losses that exceed the value at risk at confidence level $(1-\alpha)$ are those losses which are usually responsible for the disruptions of industries then we cannot ignore that value at risk is a weak point of many operational risk analyses.

Proposed solution is to use conditional value at risk which is a coherent risk measure. (Rockafellar and Stanislav Uryasev 2002) Suppose X is a random variable with a continuous distribution function F and $\alpha \in \langle 0, 1\rangle$. Conditional value at risk at confidence level $(1-\alpha)$ is defined with the following formula:

$$\mathrm{CVaR}_{1-\alpha} = \mathrm{E}\left(X | X \geq \mathrm{VaR}_{1-\alpha}\right)$$

and it can be explained as the expected value of X given that X exceeds VaR at confidence level $(1-\alpha)$.

Another formula for the conditional value at risk at confidence level $(1-\alpha)$ common in the literature (Cruz, Peters, and Shevchenko 2015) is:

$$\mathrm{CVaR}_{1-\alpha} = \frac{1}{\alpha}\int_{1-\alpha}^{1} \mathrm{VaR}_x \; dx.$$

In this research we always calculate both value at risk and conditional value at risk, thus they can be compared.

An actuarial model is used to combine the loss distribution and the frequency distribution into an aggregate loss distribution which is consequently used to calculate risk measures.

The model used for calculation of risk measures is

$$S = \sum_{n=1}^{N} X_n,$$

where the number of loss events per one year N is a random variable from a discrete frequency distribution. Numbers of breached records in one loss event $X_1, \ldots, X_n$ are random variables from a continuous loss distribution. It does not pose a problem for us that numbers of breached records are following a continuous distribution, because risk measures which interest us are usually very large numbers which can be rounded to a whole number without a change in their meaning. Number of breached records per one year S is a random variable from an aggregate loss distribution which is unknown. We usually do not know even the familiy of distributions to which the aggregate loss distribution belongs. This is caused by complicated relationships between involved random variables. The values which interest us the most are value at risk and conditional value at risk at confidence level $(1-\alpha)$ of the random variable S.

Our model based on aggregate loss distribution depends on several assumptions inspired by Cruz, Peters, and Shevchenko (2015) and A. S. Chernobai, Rachev, and Fabozzi (2007):

1. Given the number of loss events per one year N, random variables $X_1, \ldots, X_N$ are independent and identically distributed and they are

non-negative with probability one;

2. All random variables $X_1, \ldots, X_N$ are independent from the random variable N;

3. All parameters of loss and frequency distributions are known. In our case we use estimates of these parameters computed from data.

Paul Embrechts, Furrer, and Kaufmann (2003) remind another assumption that the dataset used for estimation of parameters of the model should include sufficient amount of loss events of all sizes. It means that there should be also an adequate amount of the largest loss events in the dataset. Given the unavailability of operational loss data and generally short histories of recorded losses this assumption can rarely be fully satisfied. Therefore we should interpret any research of operational risk with caution.

Because the distribution function of the aggregate loss distribution is unknown and it is complicated to get its formula analytically we use Monte Carlo simulation to estimate risk measures. If a loss distribution with estimated parameters is used to generate the numbers of breached records then the algorithm has the following steps:

1. A large number of scenarios n is chosen. For each scenario $i = 1, \ldots, n$ a number of loss events k_i is generated from the frequency distribution;

2. For each scenario $i = 1, \ldots, n$ there are k_i loss amounts generated from the loss distribution and these loss amounts are summed up. After this step there are n aggregate losses;

3. Empirical value at risk at confidence level $(1 - \alpha)$ is the $(1 - \alpha)$-th empirical quantile of the aggregate losses calculated in the previous step. Empirical conditional value at risk at confidence level $(1 - \alpha)$ is a sample mean of the $100 \cdot \alpha$ % of the largest aggregate losses calculated in the previous step.

If a generalized Pareto distribution with estimated parameters is used to generate the numbers of breached records which exceed a threshold u

and an empirical distribution function is used to generate the numbers of breached records which do not exceed the threshold u then the algorithm has the following steps:

1. A large number of scenarios n is chosen. For each scenario $i = 1, \dots, n$ a number of loss events k_i is generated from the frequency distribution;

2. For each scenario $i = 1, \dots, n$ there are k_i loss amounts generated from the loss distribution and these loss amounts are summed up. After this step there are n aggregate losses;

3. Empirical value at risk at confidence level $(1 - \alpha)$ is the $(1 - \alpha)$-th empirical quantile of the aggregate losses calculated in the previous step. Empirical conditional value at risk at confidence level $(1 - \alpha)$ is a sample mean of the $100 \cdot \alpha$ % of the largest aggregate losses calculated in the previous step.

As the number of scenarios we use 100 000. Initially, we have started with a lower number of scenarios and we have been doubling the number of scenarios until the risk measures were very similar for different values of the random seed.

6 Empirical analysis

Cyber attacks are serious threads discussed on international level. (WeLiveSecurity 2017[c]) Banks are also influenced by this thread and they need to keep adequate capital to protect against it. We conduct an empirical analysis of cyber risk similar to analysis which is expected from all banks. To analyse a part of cyber risk in a mathematical way we have chosen to evaluate the risk of data breach. The risk of data breach is only a subset of cyber risk, however it is well defined and there is no chance that we use our well established methodology without the ability to make clear conclusion.

Before we have started an empirical analysis we have chosen confidence level $\alpha = 0.05$ and high quantile for VaR and CVaR computation $\tau = 0.99$.

Furthermore, we propose these hypotheses:

1. The distribution of data breach frequency is a Poisson distribution.
2. The distribution of data breach severity is a lognormal distribution.
3. The distribution of data breach severity is a Cauchy distribution.
4. The tail of the distribution of data breach severity is a generalized Pareto distribution.

6.1 Data

"A bank's operational risk measurement system must use relevant external data (either public data and/or pooled industry data)…" (Basel Committee on Banking Supervision 2006) To test our hypotheses we use data breach dataset from Gemalto (2017) which is online publicly available. This datasets consists of 6 602 data breaches which were announced publicly between 2013 and 2016. Data breaches are divided into categories, we use observations only from categories Financial and Technology to provide results more relevant for the financial industry.

When empirical analysis is conducted there is always a danger that assumptions of the model are not satisfied. However, in this case assumptions

Distribution	Kolmogorov-Smirnov test	Anderson-Darling test
Poisson distribution	0.176498	0.00101957
Negative binomial distribution	1	0.999455

Table 1: P-values returned by goodness of fit tests for various distributions as distributions of data breach frequency

seem to be satisfied. The dataset is not a random sample, it is the whole population. Some data breaches have never been disclosed to the public and this might be a problem. There appears to be one way how we can solve it, we can consider the resulting risk measures in terms of risk that data breach happens and it is disclosed. When a data breach is not disclosed, public does not know about it and the bank can hide this data breach practically with no loss. The dataset seems to contain enough observations of large severity and thus also this assumption seems to be satisfied. By severity of data breach we mean the number of records lost in one data breach.

6.2 Results

Values of the reported risk measures are rounded to the nearest whole number.

First, we need to find the distribution of data breach frequency. We use maximum likelihood estimates to estimates parameters of two distributions and then we use goodness of fit tests. Table 1 displays p-values returned by goodness for fit tests of Poisson distribution and negative binomial distribution as distributions of data breach frequency.

Using Kolmogorov-Smirnov test we cannot reject our first hypothesis that the distribution of data breach frequency is a Poisson distribution. However, using Anderson-Darling test we can reject it. Nonetheless, negative binomial distribution seems to have a perfect fit using both any of the tests. Given how superior the fit is, we use negative binomial distribution with fitted parameters as a data breach frequency distribution in any further computations of risk measures. The fit of Poisson and negative binomial distributions is graphically represented in Figure 1 and Figure 2.

Second, we find the distribution of data breach severity using the same

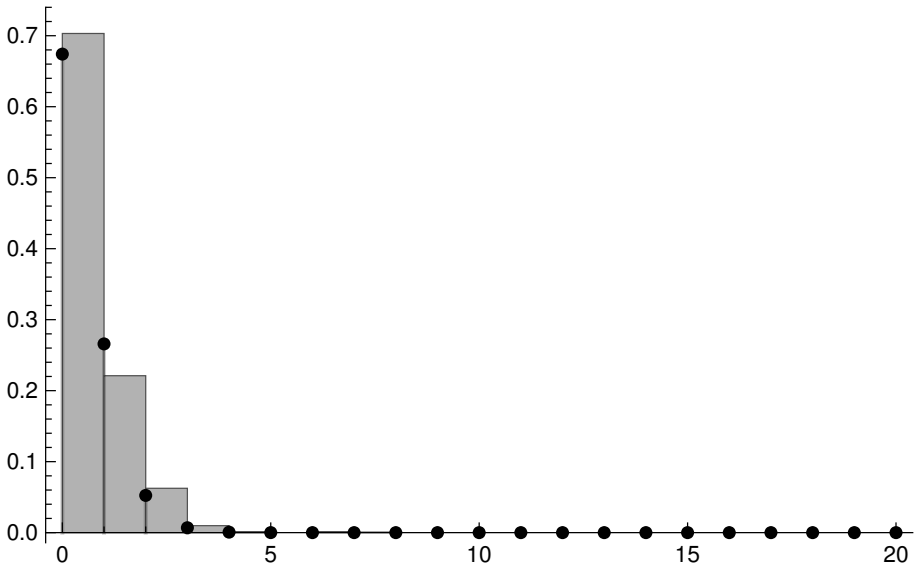

Figure 1: Histogram of the distribution of data breach frequency and fitted Poisson distribution

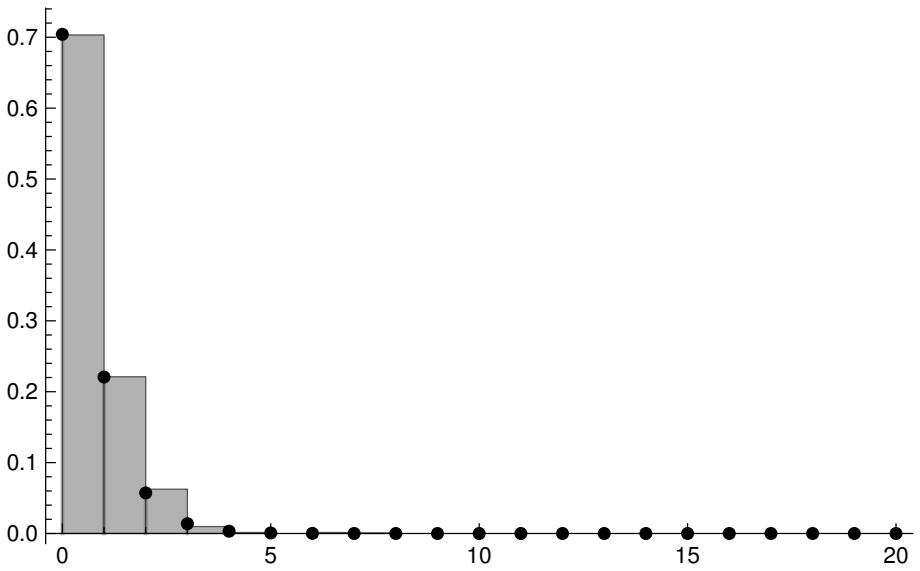

Figure 2: Histogram of the distribution of data breach frequency and fitted negative binomial distribution

Distribution	Kolmogorov-Smirnov test	Anderson-Darling test
Normal distribution	2.69318×10^{-104}	0
Exponential distribution	8.79555×10^{-288}	0
Gamma distribution	1.40911×10^{-30}	0
Lognormal distribution	0.0548474	0.136024
Weibull distribution	0.0003612	0.000017744
Pareto distribution	9.05254×10^{-35}	0
Cauchy distribution	2.87915×10^{-113}	0

Table 2: P-values returned by goodness of fit tests for various distributions as distributions of data breach severity

procedure as witch data breach frequency. Table 2 displays p-values returned by goodness of fit tests for multiple eligible distribution as distributions of data breach severity.

Using either Kolmogorov-Smirnov test or Anderson-Darling test we cannot reject our second hypothesis that the distribution of data breach severity is a lognormal distribution. However, any of these two test rejects our third hypothesis hat the distribution of data breach severity is a Cauchy distribution. Except for lognormal distribution, for any other distribution whose fit we try we can reject he null hypothesis that the distribution of data breach severity is this distribution. Figure 3 shows a histogram of data breach severity and a density function of the fitted lognormal distribution. Figure 4 shows a Q-Q plot of the fitted lognormal distribution. The picture does not show as satisfying fit as we would wish, however we rely more on the hypothesis test.

Figure 5 displays the empirical mean function of the distribution of data breach severities. Points seem to lie on a straight line for values larger than 800 000. Let us denote this value as a lower threshold. Furthermore, to be sure we use 5 000 000 as another threshold. Let us denote it as a higher threshold.

Table 3 presents p-values returned by goodness of fit tests for generalized Pareto distribution as a tail of the distribution of data breach severity. Neither Kolmogorov-Smirnov test nor Anderson-Darling test can be used to reject our

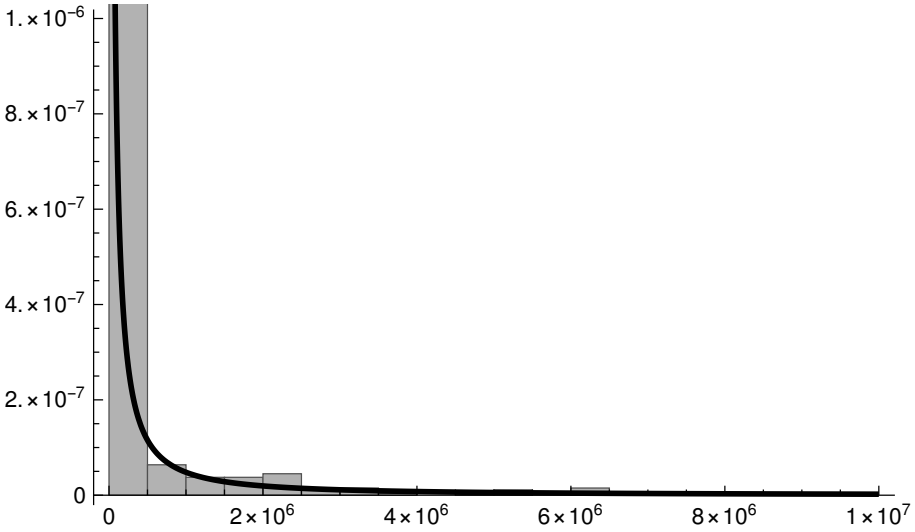

Figure 3: Histogram of the distribution of data breach severity and density function of lognormal distribution with fitted parameters

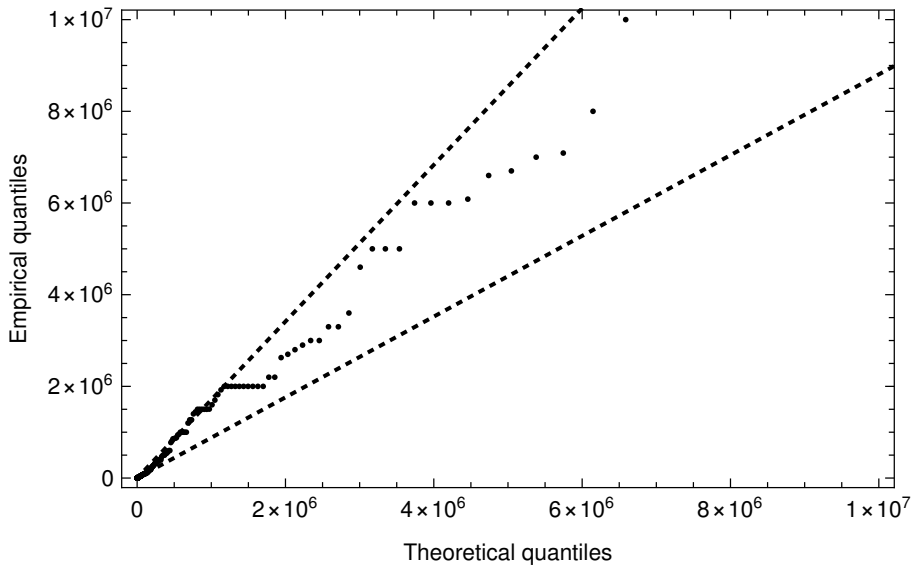

Figure 4: Q-Q plot of lognormal distribution with fitted parameters

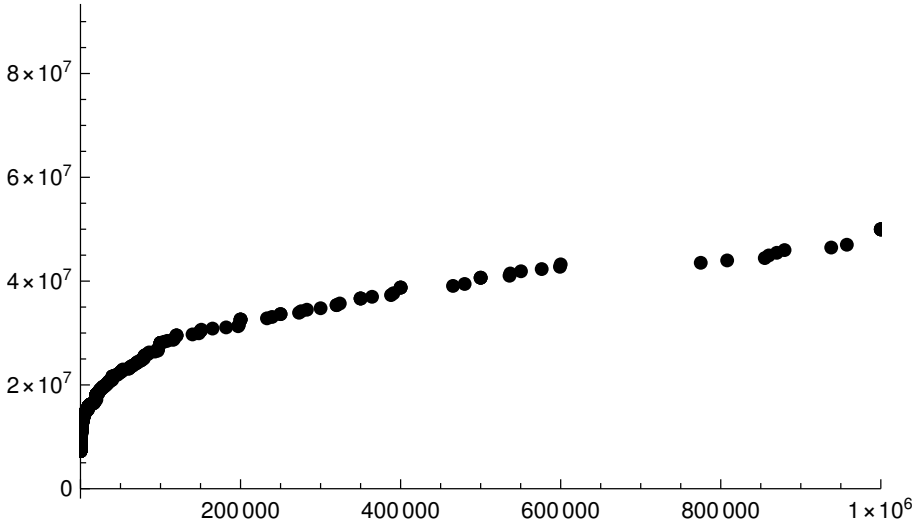

Figure 5: Empirical mean excess function of the distribution of data breach severity

Threshold	Kolmogorov-Smirnov test	Anderson-Darling test
Lower threshold	0.592211	0.793707
Higher threshold	0.846354	0.867974

Table 3: P-values returned by goodness of fit tests for generalized Pareto distribution as a tail of the distribution of data breach severity

fourth hypothesis that the tail of the distribution of data breach severity is a generalized Pareto distribution. Therefore we do not reject this hypothesis.

Figure 6 shows a histogram of the tail data breach severity distribution and a part of the density function of the fitted generalized Pareto distribution for the higher threshold. Figure 7 shows a Q-Q plot of the fitted generalized Pareto distribution for the higher threshold. Figure 8 and Figure 9 from Appendix show the same objects in a case of the lower threshold.

Table 4 presents estimates of VaR and CVaR of the aggregate loss distribution of data breaches. We report these values also for distributions whose fit to the data was not particularly fulfilling. We used fitted negative binomial distribution as the distribution of data breach frequency. We can see that if the data breach severity follows the fitted lognormal distribution then both risk measures are the largest among all other used distributions. It is impressive that under heavy-tailed distributions these risk measures are lower than under the lognormal distribution. It means that even if lognormal

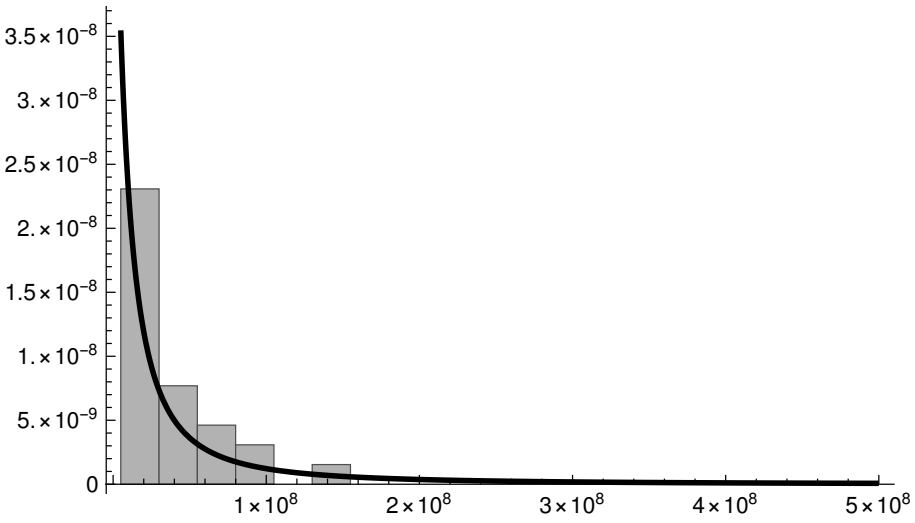

Figure 6: Histogram of the tail of the distribution of data breach severity and density function of generalized Pareto distribution with fitted parameters and higher threshold

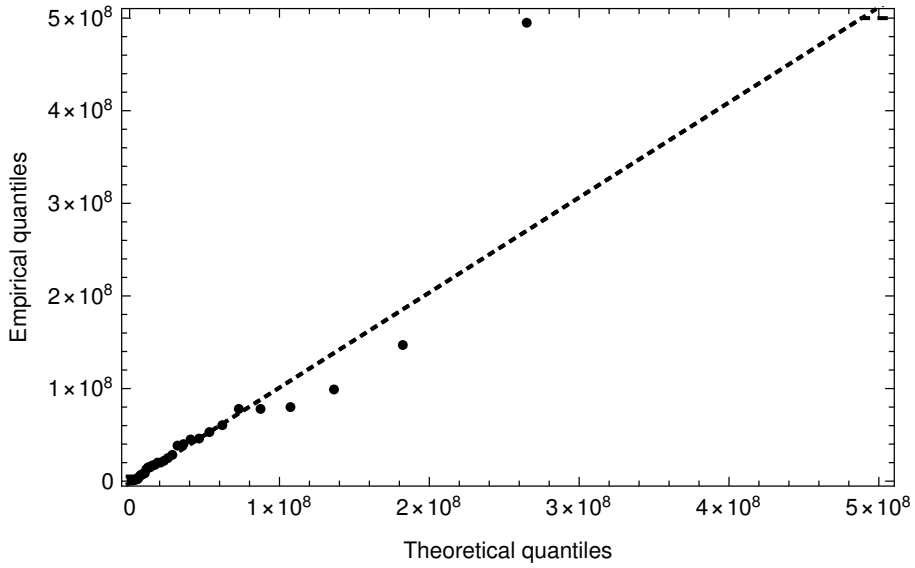

Figure 7: Q-Q plot of generalized Pareto distribution with fitted parameters and higher threshold

Distribution	VaR	CVaR
Empirical distribution	3 683 925 495	4 179 500 400
Normal distribution	3 002 254 976	3 296 890 650
Exponential distribution	1 344 312 074	1 395 854 052
Gamma distribution	1 748 197 413	1 869 738 605
Lognormal distribution	54 835 058 527	840 405 744 478
Weibull distribution	1 186 574 550	1 766 269 056
Cauchy distribution	16 497 186 950	121 257 619 936

Table 4: Values of risk measures with different severity distributions with fitted parameters.

Threshold	VaR	CVaR
Lower threshold	675 269 246 142	550 068 957 385 212
Higher threshold	29 997 814 427	660 447 391 724

Table 5: Values of risk measures when extreme value theory is used to model the severity distribution.

distribution underestimates tail of the distribution, it is not an extensive defect.

In Table 5 there are estimates of VaR and CVaR in a case when the tail of the severity distribution is approximated by generalized Pareto distribution with fitted parameters. Negative binomial distribution with fitted parameters is used in this case to approximate frequency distribution. We can see much larger number as in the previous case when we do not use extreme value theory, what is not surprising. We are particularly interested in the situation with the higher threshold, because in this situation we see better fit of the generalized Pareto distribution to the data. Fascinating finding is that under lognormal distribution and generalized Pareto distribution with the higher threshold for both risk measures we have quite close values under these two distributions.

6.3 Interpretation

It seems to be the most reasonable to prefer the extreme value theory together with the conditional value at risk over other methods we used

because generalized Pareto distribution can be used to model heavy-tailed distribution which is definitely present in our case and CVaR is a coherent risk measure with satisfactory properties. Nonetheless, lognormal distribution with the conditional value at risk looks also fairly useful.

The loss data in our dataset can be used to find distributions of frequency and severity of losses originating from the cyber risk. Statistical methods which have long history of applications to operational risk assessment are used for this task as well as for estimation of value at risk. Different from typical approaches is the interpretation of calculated value at risk. When computing value at risk (VaR), three parameters must be taken into account (A. S. Chernobai, Rachev, and Fabozzi 2007):

1. confidence level,
2. forecast horizon,
3. base currency.

Value at risk is usually calculated in a currency which is used in the country where the bank is operating. In our case it is calculated in the number of breached records, which can be just for illustration viewed as a "virtual currency".

Of course, this means that we cannot compare the estimated VaR with the assets of the bank in order to determine the capital requirement of a particular bank. However, at the same time this approach allows us to evaluate the amount of the cyber risk present in the financial industry globally. As far as we are concerned, this is much more interesting task than to estimate the capital requirement of a particular bank whose name cannot be disclosed due to the sensitivity of the loss data as it is common among several other papers on the operational risk assessment. (Lebovič 2012) This work is not purposely hiding any information about any bank or other organization. On the other hand, a discussion of the size of value at risk calculated by different techniques is provided and can be viewed as a core output of our research.

Under the assumption that the tail of the data breach severity distribution follows the generalised Pareto distribution and the data breach frequency

follows a negative binomial distribution, we can say that from a method which returns a value that is in 99 percent of cases higher than the number of breach records in financial and technology industry per one year we get value 29 997 814 427 of breached records. Similarly, under the same assumptions, a method which returns a value that is the expected value of the number of breach records in financial and technology industry per one year if this number is higher than it appears to be in 99 percent of cases we get value 660 447 391 724 of breached records.

Conclusion

Using a scenario analysis described in Rippel and Teplý (2008) we can discuss following two scenarios. According to Ponemon Institute LLC (2017) the cost of one breached record is 158 US dollars. From the dataset we use, it is possible to easily find the number of breached records per year which is over the past four years on average equal to 1.773 billion. According to Juniper Research Ltd (2017) the global cost of data breaches is 2 trillion US dollars and thus the average cost of one breached record is 1 127.57 US dollars. Under the assumption that the tail of the data breach severity distribution follows the generalised Pareto distribution and the data breach frequency follows a negative binomial distribution, the estimate of 99 percent VaR of data breach cost per one year in financial and technology industry is 4 trillion US dollars in the former scenario and 33 trillion dollars in the latter scenario.

If we compare our findings with Chalupka and Teplý (2008), we can agree that modelling of operational losses which utilizes extreme value theory performs much more satisfactory compared to situation when extreme value theory is not used. Probably it meant something when Taleb (2007) claimed: "Remember that you are a Black Swan." On the contrary, we found that lognormal distribution fits our data pretty well. We can definitely agree with Fontnouvelle et al. (2006) that external data can be used to model operational risk, or data breach risk in our case, which is nonetheless just a subset of it. It may be surprising, howeverA. S. Chernobai, Rachev, and Fabozzi (2007) provide several examples when lognormal distribution fits operational loss data. In conclusion, we are at least not alone.

The research in the empirical part of the work combines investigation of two distinct areas which are operational risk assessment and cybersecurity. We use a dataset containing numbers of records breached. Basel II Framework and corresponding literature concerned with providing examples of Advanced Measurement Approaches (Basel Committee on Banking Supervision 2006) assumes that all losses are recorded in an official monetary currency. However, Basel II Framework allows to develop custom approach to measure operational

risk and we have chosen this option.

We provided an overview of cybersecurity threads, we made recommendations about how to improve protection of banks and other organisations against cyberattacks and we provided a neat overview of operational risk assessment methodology. In the empirical part we proposed four hypothesis, three of which we did not reject. Furthermore, we calculated VaR and CVaR of data breach cost. Finally, we using external information we interpreted computed VaR in monetary terms. The major contribution of this work is that we used established methods on new data related to current topic.

References

Anderson, T. W. and D. A. Darling (1954). "A Test of Goodness of Fit". In: *Journal of the American Statistical Association* 49.268, pp. 765–769.

Angelidis, Timotheos and Stavros Degiannakis (2009). *Econometric Modeling of Value at Risk*. Nova.

Artzner, Philippe et al. (1999). "Coherent Measures of Risk". In: *Mathematical Finance* 9.3, pp. 203–228.

Balanda, Kevin P. and H. L. MacGillivray (1988). "Kurtosis: A Critical Review". In: *The American Statistician* 42.2, pp. 111–119.

Bali, Turan G. (2003). "An Extreme Value Approach to Estimating Volatility and Value at Risk". In: *The Journal of Business* 76.1, pp. 83–108.

Basel Committee on Banking Supervision (2006). *Basel II: International Convergence of Capital Measurement and Capital Standards: A Revised Framework*.

— (2016). *Standardised Measurement Approach for operational risk*.

Chalupka, Radovan and Petr Teplý (2008). *Operational Risk Management and Implications for Bank's Economic Capital – A Case Study*. IES Working Paper.

Chavez-Demoulin, V. and P. Embrechts (2004). "Advanced Extremal Models for Operational Risk". In:

Chernobai, Anna S., Svetlozar T. Rachev, and Frank J. Fabozzi (2007). *Operational Risk: A Guide to Basel II Capital Requirements, Models, and Analysis*. Wiley.

Chernobai, Anna, Philippe Jorion, and Fan Yu (2011). "The Determinants of Operational Risk in U.S. Financial Institutions". In: *The Journal of Financial and Quantitative Analysis* 46.6, pp. 1683–1725.

CNET (2017[a]). *Botnet using NSA's exploits could grow bigger than WannaCry*. URL: https://www.cnet.com/news/botnet-using-nsas-exploits-could-grow-bigger-than-wannacry/ (visited on 05/19/2017).

— (2017[b]). *Hacked NSA tools could put some Windows users at risk*. URL: https://www.cnet.com/news/hacked-nsa-tools-put-windows-users-at-possible-risk/ (visited on 05/19/2017).

— (2017[c]). *Hacked NSA tools could put some Windows users at risk*. URL: https://www.cnet.com/news/hacked-nsa-tools-put-windows-users-at-possible-risk/ (visited on 05/19/2017).

— (2017[d]). *Microsoft slams spy agencies for 'stockpiling' vulnerabilities*. URL: https://www.cnet.com/news/microsoft-slams-spy-agencies-for-stockpiling-vulnerabilities/ (visited on 05/19/2017).

CNET (2017[e]). *Russian hacker pleads guilty to get-rich-quick botnet*. URL: https://www.cnet.com/news/russian-hacker-pleads-guilty-for-his-get-rich-quick-botnet/ (visited on 05/19/2017).

— (2017[f]). *Two-factor authentication: How and why to use it*. URL: https://www.cnet.com/how-to/how-and-why-to-use-two-factor-authentication/ (visited on 05/19/2017).

— (2017[g]). *Unprecedented WannaCry attack a nightmarish 'wake-up call'*. URL: https://www.cnet.com/news/wannacry-unprecedented-ransomware-attack-a-nightmarish-wakeup-call/ (visited on 05/19/2017).

— (2017[h]). *WikiLeaks and the CIA's hacking secrets, explained*. URL: https://www.cnet.com/how-to/wikileaks-cia-hack-phone-tv-router-vault-7-year-zero-weeping-angel/ (visited on 05/19/2017).

— (2017[i]). *WikiLeaks: Apple, Google, others will get CIA hacks first*. URL: https://www.cnet.com/news/wikileaks-apple-google-samsung-cia-hacks-julian-assange/ (visited on 05/19/2017).

Cruz, Marcelo G., Gareth W. Peters, and Pavel V. Shevchenko (2015). *Fundamental Aspects of Operational Risk and Insurance Analytics: A Handbook of Operational Risk*. Wiley.

DeCarlo, Lawrence T. (1997). "On the Meaning and Use of Kurtosis". In: *Psychological Methods* 2.3, pp. 292–307.

Einmahl, John H. J., Jun Li, and Regina Y. Liu (2009). "Thresholding Events of Extreme in Simultaneous Monitoring of Multiple Risks". In: *Journal of the American Statistical Association* 104.487, pp. 982–992.

Electronic Frontier Foundation (2017). *Surveillance Self-Defense*. URL: https://ssd.eff.org/en/glossary/malicious-software (visited on 05/19/2017).

Embrechts, Paul, Hansjörg Furrer, and Roger Kaufmann (2003). "Quantifying regulatory capital for operational risk". In: *Derivatives Use, Trading & Regulation* 9.3, pp. 217–233.

Embrechts, Paul, Claudia Klüppelberg, and Thomas Mikosch (1997). *Modelling Extremal Events for Insurance and Finance*. Springer.

Fontnouvelle, Patrick de et al. (2006). "Capital and Risk: New Evidence on Implications of Large Operational Losses". In: *Journal of Money, Credit and Banking* 38.7, pp. 1819–1846.

Gemalto (2017). *The Breach Level Index*. URL: http://breachlevelindex.com/ (visited on 05/19/2017).

Ghosha, Souvik and Sidney Resnickb (2010). "A discussion on mean excess plots". In: *Stochastic Processes and their Applications* 120.8, pp. 1492–1517.

Government of the Czech Republic (2014). *Act No. 181/2014 Coll.* Act.

Juniper Research Ltd (2017). *Cybercrime will Cost Businesses Over $2 Trillion by 2019*. URL: https://www.juniperresearch.com/press/press-releases/cybercrime-cost-businesses-over-2trillion (visited on 05/19/2017).

Kemp, Malcolm (2011). *Extreme Events: Robust Portfolio Construction in the Presence of Fat Tails*. Wiley.

Kulich, Michal (2014). *Přehledový větník: NMFM301 Statistika pro finanční matematiky*.

Lab, Kaspersky (2017[a]). *7 questions about 11-11, answered*. URL: https://blog.kaspersky.com/kaspersky-os-7-facts/14084/ (visited on 05/19/2017).

— (2017[b]). *Lazarus: Modus operandi and countermeasures*. URL: https://blog.kaspersky.com/lazarus-modus-operandi-and-countermeasures/6716/ (visited on 05/19/2017).

— (2017[c]). *The greatest heist of the century: hackers stole $1 bln*. URL: https://blog.kaspersky.com/billion-dollar-apt-carbanak/7519/ (visited on 05/19/2017).

Lebovič, Michal (2012). *The Use of Coherent Risk Measures in Operational Risk Modeling*. Diploma Thesis.

Mejstřík, Michal, Magda Pečená, and Petr Teplý (2015). *Bankovnictví v teorii a praxi / Banking in Theory and Practice*. Charles University in Prague, Karolinum Press.

Nair, Jayakrishnan, Adam Wierman, and Bert Zwart (2013). *The Fundamentals of Heavy Tails: Properties, Emergence, & Identification*.

Ponemon Institute LLC (2017). *2016 Cost of Data Breach Study: Global Analysis*. URL: https://www.ibm.com/security/data-breach/ (visited on 05/19/2017).

Rachev, Svetlozar T., Young Shin Kim, and Michele L. Bianchi (2011). *Financial Models with Levy Processes and Volatility Clustering*. Wiley.

Rippel, Milan and Petr Teplý (2008). *Operational Risk – Scenario Analysis*. IES Working Paper.

Rockafellar, R. Tyrrell and Stanislav Uryasev (2002). "Conditional value-at-risk for general loss distributions". In: *Journal of Banking & Finance* 26.7, pp. 1443–1471.

Sarykalin, Sergey, Gaia Serraino, and Stan Uryasev (2008). *Value-at-Risk vs. Conditional Value-at-Risk in Risk Management and Optimization*. Tutorials in Operations Research, INFORMS 2008.

Taleb, Nassim Nicholas (2007). *The Black Swan: The Impact of the Highly Improbable*. Random House.

WeLiveSecurity (2017[a]). *10 things to know about the October 21 IoT DDoS attacks*. URL: https://www.welivesecurity.com/2016/10/24/10-things-know-october-21-iot-ddos-attacks/ (visited on 05/19/2017).

— (2017[b]). *False positives can be more costly than a malware infection*. URL: https://www.welivesecurity.com/2017/05/09/false-positives-can-costly-malware-infection/ (visited on 05/19/2017).

WeLiveSecurity (2017[c]). *Financial cybersecurity 'needs to be a key agenda item at G20'*. URL: https://www.welivesecurity.com/2016/09/01/financial-cybersecurity-needs-key-agenda-item-g20/ (visited on 05/19/2017).

— (2017[d]). *Five interesting facts about the Morris worm (for its 25th anniversary)*. URL: https://www.welivesecurity.com/2013/11/06/five-interesting-facts-about-the-morris-worm-for-its-25th-anniversary/ (visited on 05/19/2017).

— (2017[e]). *Google announces 'Vulnerability Research Grants'*. URL: https://www.welivesecurity.com/2015/02/02/google-announces-vulnerability-research-grants/ (visited on 05/19/2017).

— (2017[f]). *Huge ransomware outbreak disrupts IT systems worldwide: WannaCryptor to blame*. URL: https://www.welivesecurity.com/2017/05/13/wanna-cryptor-ransomware-outbreak/ (visited on 05/19/2017).

— (2017[g]). *Microsoft issues warning after Xbox Live certificate 'inadvertently' leaks*. URL: https://www.welivesecurity.com/2015/12/11/microsoft-issues-warning-xbox-live-certificate-inadvertently-leaks/ (visited on 05/19/2017).

— (2017[h]). *Microsoft says patch your Windows PCs now against critical security vulnerabilities*. URL: https://www.welivesecurity.com/2016/08/10/microsoft-says-patch-windows-pcs-now-critical-security-vulnerabilities/ (visited on 05/19/2017).

— (2017[i]). *Wi-Fi security – routers "like fish in a barrel"*. URL: https://www.welivesecurity.com/2014/08/13/wi-fi-security-routers-like-fish-in-barrel/ (visited on 05/19/2017).

Westfall, Peter H. (2014). "Kurtosis as Peakedness, 1905-2014. RIP". In: *The American Statistician* 68.3, pp. 191–195.

List of tables and figures

List of Figures

List of Tables

Appendix

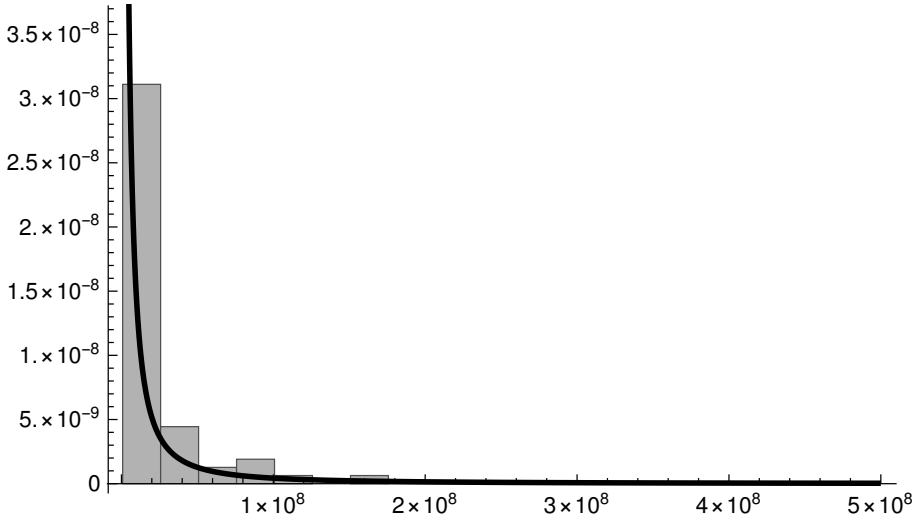

Figure 8: Histogram of the tail of the distribution of data breach severity and density function of generalized Pareto distribution with fitted parameters and lower threshold

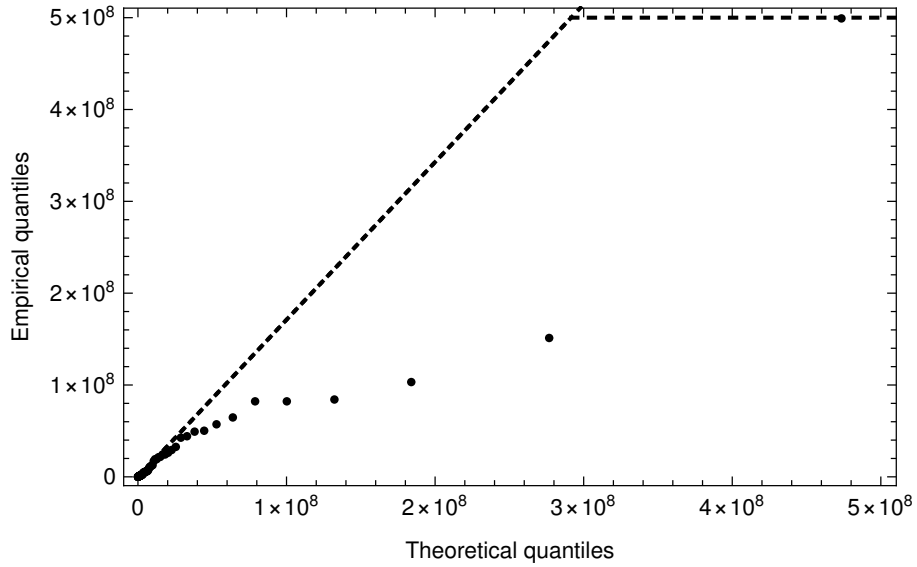

Figure 9: Q-Q plot of generalized Pareto distribution with fitted parameters and lower threshold