CHARLES UNIVERSITY IN PRAGUE
FACULTY OF SOCIAL SCIENCES
Institute of Economic Studies

Michal Lebovič

The Use of Coherent Risk Measures in Operational Risk Modeling

DIPLOMA THESIS

Prague 2012

Bibliographical Record

LEBOVIČ, M. (2012): *Use of Coherent Risk Measures in Operational Risk Modeling*. Diploma Thesis, Charles University in Prague, Faculty of Social Sciences, Institute of Economic Studies, 2012, pages: 112. Thesis supervisor: PhDr. Petr Teplý, Ph.D.

Author: **Michal Lebovič**

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

Academic Year: **2011/2012**

ABSTRACT

The debate on quantitative operational risk modeling has only started at the beginning of the last decade and the best-practices are still far from being established. Estimation of capital requirements for operational risk under Advanced Measurement Approaches of Basel II is critically dependent on the choice of risk measure, which quantifies the risk exposure based on the underlying simulated distribution of losses. Despite its well-known caveats Value-at-Risk remains a predominant risk measure used in the context of operational risk management. We describe several serious drawbacks of Value-at-Risk and explain why it can possibly lead to misleading conclusions. As a remedy we suggest the use of coherent risk measures – and namely the statistic known as Expected Shortfall – as a suitable alternative or complement for quantification of operational risk exposure. We demonstrate that application of Expected Shortfall in operational loss modeling is feasible and produces reasonable and consistent results. We also consider a variety of statistical techniques for modeling of underlying loss distribution and evaluate extreme value theory framework as the most suitable for this purpose. Using stress tests we further compare the robustness and consistency of selected models and their implied risk capital estimates calculated with VaR and ES statistics.

KEYWORDS: operational risk, risk measures, value-at-risk, expected shortfall, advanced measurement approach, loss distribution approach, extreme value theory

JEL CLASSIFICATION: C15, G21, G32

ABSTRAKT

Debata o metodách modelování operačních rizik byla otevřena teprve počátkem minulého desetiletí a je dodnes živá a plná otevřených otázek. Pro pokročilé metody měření (AMA) v rámci pravidel Basel II je odhad kapitálové přiměřenosti vůči operačnímu riziku zásadně ovlivněn zvolenou metodou měření rizika, která je použita k výpočtu peněžního vyjádření expozice vůči riziku ze simulovaného statistického rozdělení operačních ztrát. Nejčastěji používanou metodou měření rizika je Value-at-Risk (ohrožená hodnota), přestože je v odborné literatuře považována za překonanou. V této práci popisujeme nejvážnější problémy spjaté s metodou VaR a vysvětlujeme, proč může v kontextu modelování operačních rizik vést k zavádějícím závěrům. Jako vhodnou alternativu, či alespoň doplněk, k ohrožené hodnotě pro účely výpočtu kapitálové přiměřenosti potom navrhujeme třídu metod měření rizika známou pod přízviskem koherentní. Konkrétně vybíráme koherentní metodu zvanou Expected Shortfall a v empirické části našeho výzkumu ukazujeme, že produkuje smysluplné a konzistentní výsledky a že je v daném kontextu velmi užitečným nástrojem. Důležitou rovinou práce je ovšem i samotné modelování distribuce operačních ztrát. K tomu využíváme celou řadu statistických metod, přičemž se ukazuje jako nejvhodnější technika teorie extrémních hodnot. Pomocí zátěžových testů dále testujeme robustnost a konzistenci vybraných modelů ztrát i výpočtů kapitálové přiměřenosti daných oběma zmíněnými metodami měření rizika.

KLÍČOVÁ SLOVA: operační riziko, metody měření rizika, ohrožená hodnota, expected shortfall, pokročilé metody měření (AMA), metoda distribuce ztrát (LDA), teorie extrémních hodnot

KLASIFIKACE JEL: C15, G21, G32

ACKNOWLEDGEMENT

Hereby I would like to express my gratitude to PhDr. Petr Teplý Ph.D. for supervising my thesis and for his invaluable comments and suggestions.

DECLARATION

I hereby declare that this thesis is the result of my own independent work and that I used solely the listed literature and resources.

Prague, May 15, 2012 Michal Lebovič

Master Thesis Proposal

Institute of Economic Studies
Faculty of Social Sciences
Charles University in Prague

Author:	**Bc. Michal Lebovič**	Supervisor:	PhDr. Petr Teplý Ph.D.
E-mail:	lebovicm@gmail.com	E-mail:	teply@fsv.cuni.cz
Phone:	+420 724 593 212	Phone:	+420 222 112 320
Specialization:	Finance, Fin. Markets and Banking	Defense Planned:	May 2012

Proposed Topic:

The use of coherent risk measures in operational risk modeling

Topic Characteristics:

Since its inclusion in the Basel capital directives in 2001 the operational risk became widely acknowledged as one of the most pervasive types of risk faced by the financial institutions. Supported by Basel II policy guidelines Value-at-Risk has quickly become a dominant quantitative tool for operational risk measurement, just as it did for other classes of risk. However many authors have been drawing attention to the inherent weaknesses of VaR and stressed the need to look for feasible alternatives. Credibility of VaR also greatly suffered during the recent financial crisis because the fatal mismanagement of risk was recognized as one its important contributing factors. This thesis is intended as a contribution to the crucial discussion on the appropriateness of various methods for operational risk modeling and management.

There are two main issues to be examined in this thesis. First it will critically evaluate VaR and try to assess the limits for its reasonable use as an operational risk modeling and management tool. That will be done by both exploring weaknesses in the mathematical and statistical properties of VaR and by analyzing its practical use (or misuse) prior and during the financial crisis. Second, it will introduce and analyze an alternative to VaR class models - the group of models known as „Coherent risk measures“ as defined in Artzner et al. (1999). I will discuss in details the four mathematical properties that a risk measure must satisfy in order to be called „coherent“ - translation invariance, subadditivity, positive homogeneity and monotonicity - and show why they define a set of necessary “good qualities” a risk measure should have and why they form a logical basis for operational risk modeling. I will also introduce several specific coherent risk measures such as SPAN or Expected shortfall methods and discuss their usefulness for the operational risk modeling. Critical comparison of these methods to VaR is the focal point of this thesis as the primary concern is whether they represent a significant improvement over VaR.

Most importantly the discussion will be supplemented with an empirical analysis of a dataset of historical operational losses of a large Czech commercial bank. Using all of the described risk modeling techniques I will assess their respective estimates of required capital charges along with the robustness analysis and use the results to draw conclusions about the appropriateness of the employed methods.

Hypotheses:

1. Value-at-Risk is unable to consistently provide reasonable estimates of adequate risk capital charges or there are some general scenarios where it is unable to do so.
2. Extensive use of VaR as a key risk modeling method contributed to the fatal mismanagement of risk and was one of the contributing factors to the financial crisis.
3. The most commonly used versions of VaR do not satisfy the general properties of coherent risk measures.
4. Coherent risk measures provide better estimates of adequate risk capital charge than VaR and provide a reasonable basis for the operational risk measurement under the Basel regulatory Framework.

Methodology:

Methods use throughout the thesis:
- Comparative analysis of VaR and Coherent risk measures
- Axiomatic derivation of coherent risk measures and its properties.
- Mathematical statistics used to describe risk measurement methods
- Empirical data analysis using the statistical packages

Outline:

1. Introduction
2. Basics of operational risk modeling
3. Value-at-Risk and operational risk modeling: critical analysis
4. Coherent risk measures
    a. Definitions, properties and methodological issues for operational risk modeling
    b. Expected Shortfall, SPAN and other coherent risk measures
5. VaR and Coherent risk measures: empirical analysis
    a. Methodology
    b. Results and robustness checks
6. Conclusions

Core Bibliography:

- Artzner, P., Delbaen, F., Eber, J., and Heath, D. (1999). „*Coherent Measures of Risk*," Mathematical Finance 9, pp. 203-228.
- Artzner, P., Delbaen, F., Eber, J., and Heath, D. (1997). „*Thinking Coherently*", RISK 10, pp. 68-71.
- BIS (1999). „*A New Capital Adequacy Framework*," http://www.bis.org.
- BIS (2010). „*Operational Risk – Supervisory Guidelines for the Advanced Measurement Approaches*," http://www.bis.org.
- BIS (2010). „*Sound Practices for the Management and Supervision of Operational Risk*," http://www.bis.org.
- Chernobai, A., Fabozzi, F., and Rachev, S. (2007). „*Operational Risk: A Guide to Basel II Capital Requirements, Models, and Analysis,*" John Wiley & Sons, New Jersey.
- Cruz, M. G. (2002). „*Modeling, Measuring and Hedging Operational Risk*," John Wiley & Sons, New York.
- Jorion, P. (2000). "*Value-at-Risk: The New Benchmark for Managing Financial Risks,*" 2nd ed., McGraw-Hill, New York.
- Koji, I., Masaaki K. (2005). „*On the significance of expected shortfall as a coherent risk measure*," Journal of Banking & Finance, 29, Risk Measurement, pp. 853-864.
- Yamai Y., and Yoshiba, T. (2002). „*Comparative Analysis of Expected Shortfall and Value-at-Risk: Their Estimation, Decomposition, and Optimization*," Monetary and Economic Studies, January, pp. 87-121.
- Yamai Y., and Yoshiba, T. (2002). „*On the Validity of Value-at-Risk: Comparatie Analysis with Expected Shortfall*," Monetary and Economic Studies, January, pp. 57-85.

Michal Lebovič — Author

PhDr. Petr Teplý Ph.D. — Supervisor

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Operational risk management is certainly the youngest of risk management disciplines. Growing recognition of operational risk as a standalone risk class culminated in its inclusion in Basel banking regulation rules in 2001 and dismantled the idea of operational risk as a "residual to market and credit risk" for good. Following a definition from Basel II we regard operational risk as "the risk of direct or indirect loss resulting from inadequate or failed internal processes, people and systems or from external events" (BIS (2001b)). This thesis is intended as a contribution to the crucial discussion on the appropriateness of certain quantitative techniques for operational risk modeling at this apt time, when the discussion on the topic is at its highest and best-practices are far from being established.

Particular question of interest is how to quantify the operational risk exposure from the modeled distribution of operational losses over certain period. A predominant tool used for this purpose is Value-at-Risk. Yet we stress that Value-at-Risk is just one of many available risk measures and appealing alternatives exist for the problem at hand - the main aim of our research is to explore the applicability and usefulness of a class of risk measures known as "coherent risk measures" for an operational risk modeling.

Search for alternatives to Value-at-Risk in the context of quantitative operational risk management is motivated by a simple paradox. The distinguishing feature of operational risk data is that it is dominated by events of very low frequency but huge magnitude; most authors would agree that capturing the behavior of such events is the cornerstone of operational risk modeling. Yet for the description of these extreme events the Value-at-Risk measure is used which answers the question "what is our minimum loss in case the extreme event occurs?" With a little exaggeration one could say that VaR discards such information from the loss distribution that is the most telling of the nature of risks in question.

Concept of coherence was introduced in famous articles of Artzner et al. (1997, 1999) and later extended by Delbaen (2002). It was the first attempt to axiomatize the debate on the risk measures and provoked a huge number of responses in the academic literature.

The basic contribution of the article was to characterize coherent risk measures either as a minimum amount of capital to be added to the payoff of random risk variable X to achieve an acceptable position, or as the supremum of the expected loss over a set of so-called *generalized scenarios.* It also led to a creation of the most prominent of coherent risk measures - Expected Shortfall. We consider Expected Shortfall (sometimes also known as Conditional Value-at-Risk) as a possible answer to the above-mentioned paradox and we propose it as an alternative or a supplement to value-at-risk for the purposes of operational risk modeling.

For continuous distributions expected shortfall coincides with another conceptually simple risk measure - known as Conditional Tail Expectation, Mean Excess Loss or Tail Value-at-Risk - defined as the expected loss under the condition that it exceeds VaR threshold (Hurliman (2003)). However such risk measure is not coherent for the discrete distributions, hence also for most of the applications in operational risk modeling where the distribution of losses is typically modeled numerically using a Monte Carlo simulation engine. In contrast, Expected Shortfall satisfies the properties of coherence under general distributions (first proved by Pflug (2000)) and it is defined as a weighted average of value-at-risk and losses strictly exceeding VaR (e.g. Rockafellar and Uryasev (2000)).

Both Value-at-Risk and Expected shortfall are percentile-based risk measures which are suitable for precise estimation of extreme right quantiles of loss distributions. Even though VaR is still predominant amongst practitioners ES has theoretical properties superior to VaR and its popularity has been growing in last decade. For many risk management applications it is now considered a serious alternative to VaR (e.g. Dowd (2006), Cizek et al. (2011) or Krokhmal et al. (2001)). Even though it has been only slowly finding its way into operational risk literature, alternative risk measures that would capture the tail risk better than VaR have been called for (e.g. Moscadelli (2004)) and ES has already appeared in several recent empirical applications such as Lee and Fang (2010) or Giacometti et al. (2008).

But application of any risk measure would be meaningless without a reasonable model of the distribution of aggregate operational losses. Finding a proper model of operational losses is an inherent part of our research and we employ a variety of statistical techniques pursuing this goal. In this regard we focus on so-called Loss Distribution Approach methodology. We try to fit a wide list of statistical distributions to our operational loss data, such as lognormal, Weibull, gamma, or a less notorious generalized 4-parameter g-and-h distribution. We also employ Extreme Value Theory that is designed specifically for the precise modeling of extreme quantiles as we find that conventional statistical modeling does not produce satisfactory results. Hence a second contribution of this thesis is embodied in our conclusions on the suitability of various loss modeling techniques for the purposes of operational risk modeling. Technically we broadly follow the guidelines given in Chernobai et al. (2007) and compare our results with the most prominent empirical research on operational risk, such as Moscadelli (2004), de Fontnouvelle (2005) or Jobst (2007).

To recapitulate the presented topic, our hypotheses can be summarized as follows (we do not state the alternatives as they are somewhat obvious):

$\mathrm{H}_0^{\mathrm{A}}$: *Extreme Value Theory provides a suitable framework for quantitative operational risk management and produces reasonable and consistent models of operational risk exposure.*

$\mathrm{H}_0^{\mathrm{B}}$: *Expected Shortfall is an applicable and consistent tool for estimation of operational risk capital charges under variety of statistical modeling methods and can be used as a suitable complement or alternative to Value-at-Risk.*

The remainder of the thesis has a following structure. In first chapter we give a basic explanation of operational risk and define some basic concepts necessary for understanding of the operational risk modeling. Chapter 2 describes specific quantitative methods that are used in this research. Chapter 3 concerns the use of different risk measures as one of the distinct issues of operational risk modeling; it explains the concept of coherence and gives rationale for the use of coherent risk measures as an efficient alternative or supplement to Value-at-Risk. Finally in Chapter 4 we present the results of our empirical research using real operational risk data where we apply the theoretical findings from previous chapters to search for the answers on hypothesis stated above.

Chapter 1. Operational risk – basic concepts

1.1. What is operational risk?

Even though certain considerations and managerial practices implied by operational risk have been an inevitable part of manager's toolbox for many decades operational risk was only really institutionalized by the New Basel Capital Accord, commonly known as Basel II, established in 1999 (see BIS, 2001a, for a revised version of the original document). It was this document that gave operational risk a fair status amongst other kinds of risk and recognized it as one of the crucial risk factors faced by the financial institutions. The regulatory capital charges against operational risk have been included in Basel II in one of the amendments from 2006 (see e.g. BIS, 2010a). Broader description and the fascinating history of the operational risk can be reviewed in more detail in Power (2005), BIS (2001b, 2010), Cruz (2002) or Chernobai et al. (2007).

Widely accepted definition[1] of operational risk as formulated by Basel Committee on Banking Supervision in BIS (2001b) is stated below:

> *"Operational risk is the risk of direct or indirect loss resulting from inadequate or failed internal processes, people and systems or from external events."*

Basel Committee further distinguishes 7 operational risk event types (Table 1.1). It also stresses the need to treat the operational risk independently for various business lines within the bank (list of business lines can be found in Table 1.2 in section 1.5). This classification (details can be found e.g. in BIS (2006)) has been accepted as a standard and it is used in empirical literature where the analysis of various business line effects or event

[1] Goodheart (2001) reviews definitional issues on operational risk pointing out that the definition "was clarified to exclude reputational and strategic risk and focuses on the causes of loss." Also notice that it unambiguously defines this risk as downside loss, not as variation about expected mean.

type effects is considered (for example in Moscadelli (2003), de Fontnouvelle et al. (2003) or Dutta and Perry (2007)).

Table 1.1: Event type classes in operational risk

Operational risk events classification	
Event type	***Examples***
1. Internal fraud	Theft and fraud, unauthorized trading
2. External fraud	Theft and fraud, software security breach
3. Employment practices and workplace safety	Employee relations, safety of environment
4. Clients, products and business practices	Improper business or market practices, product flaws
5. Damage to physical assets	Natural disasters
6. Business disruption and system failures	Software or hardware failures
7. Execution, delivery, process management	Incorrect transaction execution, model failures

Source: Author based on BIS (2006).

1.2. Note on the nature of operational risk management

As one of opening notes, we need to emphasize that in the paradigm of this thesis operational risk management is a quantitative exercise where risk is regarded as an expected statistical distribution of operational losses. But we acknowledge that it is only one (and perhaps not the most important) of the dimensions of this discipline. By itself it in no way helps neither to mitigate operational risk events in the bank nor reduce the resulting losses, which should be of primary concern to risk managers. That can only be done by careful configuration and control of internal and external policies, systems, processes and other institutional factors within the company (for discussions on this dimension of operational risk management reader can view Power (2005) and references therein). Nonetheless the quantitative aspect of operational modeling is still extremely important for at least two reasons. First, it is necessary to accommodate the regulatory requirements of the Basel II rules. Second reason is purely practical and stems from the concept of economic capital. Even if it was not required by the regulator, the bank - if it wanted to remain in business - would need to have some quantitative apprehension of future losses in order to create corresponding capital reserves as a protection against insolvency in case of occurrence of large unexpected losses. Method we present in this thesis is an attempt to build such apprehension in a best possible way based on the past experience of the bank with operational risks.

1.3. Quantitative concepts of risk and risk measures

This section gives only a brief and rather non-technical summary of the mutually related concepts of risk[2] and risk measures that the reader needs to comprehend before proceeding with reading the text. From the viewpoint of risk quantification and measurement, risk can be regarded as a random variable X representing random profit or loss of a given position. It can be defined in two ways – either as two-sided, i.e. as variation about expected mean where both positive and negative values can be attained, or as one-sided with non-positive values only. The first case is the domain of market risk for instance. When holding certain financial instrument one can expect its value to either decrease or increase during the holding period, which is equivalent to undertaking a two-sided risk where both profits and losses are possible. For some other classes of risks – operational risk amongst them – the other approach is necessary. Operational risk events are always defined as down-side losses, no matter which exact definition is used. Hence it will be necessary to adjust some of the risk literature concepts to suit the context of operational risk, because most of them are elaborated under the assumption of two-sided risk. Furthermore in operational risk the losses are typically regarded as positive numbers (the higher the value, the higher the loss), hence the highest losses are associated with the extreme right quantiles of the loss distribution.

Given the random variable X representing risks, quantification of risk exposure is done using risk measures, mathematical rules mapping X to the real numbers – usually in some currency units – in some meaningful way. Risk measures hence allow us to translate the known (estimated) statistical distribution of losses into a single comprehensible and comparable statistic. By far the most common risk measure is Value-at-Risk which is a simple α-quantile of the distribution of losses. Although many are not even aware of this fact, some alternatives to VaR have been developed over the last two decades – one of those alternatives is central to this thesis and is known as "coherent risk measures" (whole Chapter 3 is devoted to a more rigorous description and discussion of VaR and coherent risk measures).

[2] The explanation given in this section is not intended as a general definition of the term "risk", which would be an intricate task. We are only explaining the concept as it is used in the context of quantitative modeling.

1.4. Regulatory and Economic Capital

Banks keep certain amount of risk capital in order to form a protective buffer against the losses caused by various classes of risks. In the next section when we write about Basel II we are referring to the concept of regulatory capital. Basically it is an amount of capital reserves a bank must hold in order to satisfy the demands of the regulator and that is the amount set by Basel II capital adequacy rules. We must further differentiate between the concepts of minimum capital requirement, which is a minimum value of a ratio of amount of capital[3] in the bank and its risk-weighted assets, and the regulatory capital charge against operational or other risk which is a minimum amount of capital that the bank must keep for the specific purpose of protecting itself against the losses caused by the given type of risk. Next section gives some first insight on how the capital charges are determined - but in fact that is a paramount question and it is interconnected with all of the issues discussed in this thesis.

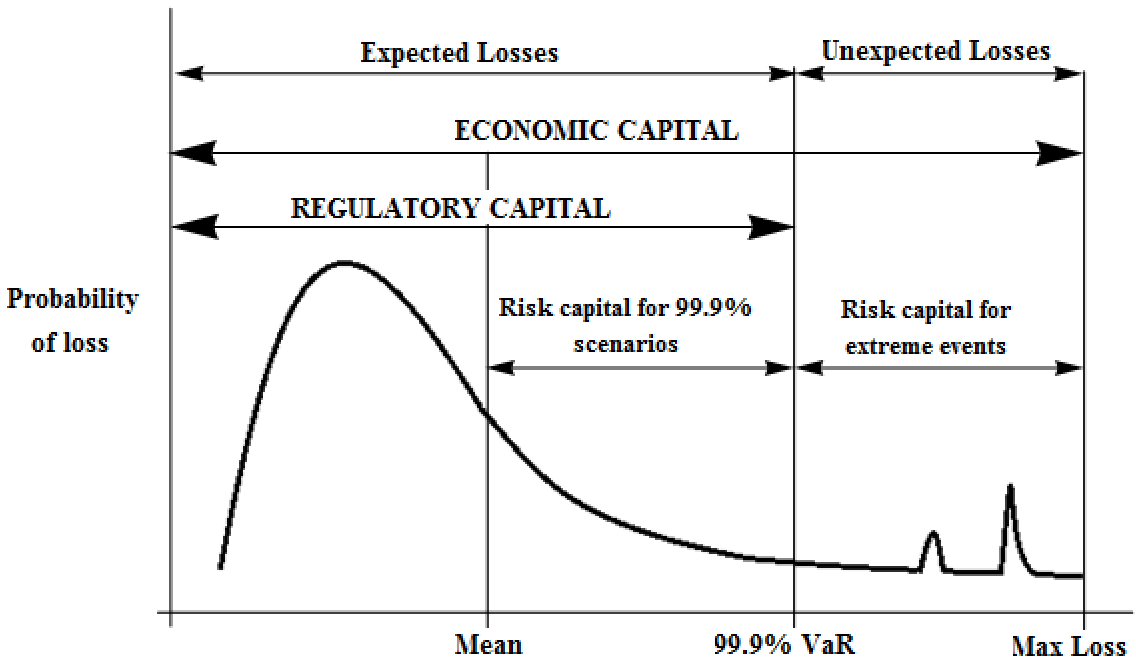
A different concept that must be distinguished from the concept of regulatory capital is the economic capital. Chernobai et al. (2007) defines it as “the amount of capital market forces dictate for risk in a bank.” Another possible definition is provided by Mejstrik et al. (2007): “(economic capital) is a buffer against future unexpected losses brought about by credit, market, and operational risks inherent in the business of lending money.” Apt distinction between the two kinds of capital is also due to Chorafas (2006, p.61) who writes that regulatory capital is “minimum amount needed to have a license (and) economic capital is the mean amount necessary to be in business…” Same author also stresses that regulatory capital corresponds mostly to expected losses which are of higher frequency but lower impact while economic capital (p.63) “can extend all the way to extreme events” and defines it as a cushion for unexpected losses - yet we should note here that for operational risk this distinction is not so clear as under Advanced Measurement Approaches (see section 1.5) defined in Basel II the regulatory capital charge is supposed to cover also most, but not all, of the unexpected losses (in fact it should cover all operational losses with 99.9% confidence level).

Even though the definitions are somewhat different (and surely one can encounter many others in the literature) they all have one common feature that distinguish economic

[3] Not all kinds of capital qualify to enter this ratio though. See Chernobai et al. (2007) or Chorafas (2006) for detailed classification of different kinds of capital and their eligibility for Basel II capital adequacy.

and regulatory capital - the latter is imposed by the regulator while the former is a bank's self-imposed prudency measure, a managerial, not regulatory requirement. Amount of economic capital to be held by the bank is a matter of managerial discretion but it can never be less than the regulatory capital - the concepts are tied by the fact that regulatory capital forms an infimum of a capital reserve the bank can rationally and voluntarily hold while economic capital is the supremum. That is why it is necessary to have the regulatory requirements set sensibly – amounts too high unnecessarily tie up capital that could be used elsewhere and might hamper bank's business while amounts too low provide leeway[4] for banks to take excessive risks without proper hedging.

Figure 1.1: Regulatory and economic capital of the bank



Source – Author based on Chorafas (2006)

1.5. Basel II and operational risk

Basel II Capital Accord lies in the heart of the international banking regulation system and hence it is one of the key influences on the risk management practices adopted worldwide (apart from being an important mediator of theoretical discourse). For this

[4] One could controvert whether the stringent regulatory requirements imposed on prudential behavior of banks are even necessary, but that would be a topic for another thesis.

reason it is necessary to offer here a short summary of Basel II rules imposed on the measurement, management and reporting of operational risk.

Basel II adopted a three-pillar structure where each pillar is designed to perform a specific role, all three together meant to provide a necessary minimum regulation for reinforcing the stability of international banking systems and financial markets. Main concern of Basel regulation rules is to ensure the proper management of credit risk (since 1988), market risk (since 1996) and operational risk (since 2001), management of other risks is left at the discretion of individual banks. Basic description of the three pillars follows (see BIS, 2001a, for details):

a) *Pillar I – Minimum capital requirements.*

   First pillar sets the rules for quantification of various sources of risk and the minimum regulatory capital charge required to offset them. It defines three different admissible approaches to the measurement of operational risk capital charge (they are described in the next subsection).

b) *Pillar II – Supervisory review of capital adequacy*

   Second pillar generally focuses on the quality and compliance of used risk management techniques and internal controls.

c) *Pillar III – Market discipline & public disclosure*

   Last pillar focuses on the transparency of risk management; it comprises rules and requirements on the disclosure activities of banks, reporting to shareholders and customers etc.

For the purposes of this thesis the first pillar is critical because it defines the rules for the calculation of required regulatory capital charge against operational risk and thereby sets the technical standards of quantitative operational risk management. Rest of this section provides a more in-depth overview of the techniques of capital charge estimation admissible under Basel II regulatory framework.

Computing operational risk capital charge: the three approaches

Pillar I of Basel regulation system defines three general approaches to the operational capital charge measurement:

(i) Basic Indicator Approach

(ii) Standardized Approach

(iii) Advanced Measurement Approach

Using the terminology of Chernobai et al. (2007), Basic Indicator Approach (BIA) and Standardized Approach (SA) are called "*top-down*" as they determine capital charge as certain fixed proportion of income, hence they proxy the operational risk exposure using solely macro-level indicators disregarding the actual causes of loss events. Advanced measurement approaches (AMA) are *bottom-up* as they are based on internal database of historical losses of individual banks. Hence they are supposed to reflect more closely the risk profile of individual banks than bottom-up approaches.

Table 1.2: Overview of Basic Indicator Approach and Standardized Approach to operational risk capital charge assessment.

Basic Indicator Approach:

$$K_{BIA} = \alpha \frac{\sum_{t=1}^{3} max\{GI_t; 0\}}{n}$$

Standardized Approach:

$$K_{SA} = \frac{\sum_{t=1}^{3} max\{\sum_{i=1}^{8} GI_{t,i}\beta_i ; 0\}}{3}$$

Values of β under Standardized Approach for various business lines:

Business line	β
1. Corporate finance	18%
2. Trading and sales	18%
3. Retail banking	12%
4. Commercial banking	15%
5. Payment and settlement	18%
6. Agency services	15%
7. Asset management	12%
8. Retail brokerage	12%

Source: Author based on BIS (2006)

Let us briefly explain the "top-down" approaches inbuilt in Basel regulation system (overview is in Table 1.2). For BIA which is the simplest of the three approaches the capital charge is set trivially as a 15% of average annual positive gross income (GI) over past three years (*n* in the formula in Table 1.2 is the number of years with positive gross income amongst past three years). Charge (α) was set to 15% most likely because in time of the discussions of Basel II amendments this value was an industry-wide average of economic capital charge for operational losses. In a sense Basel II only forced the banks to do what they were already doing (at least on average). A little more complicated approach is SA – it is essentially based on the same idea with the difference that the banks' activities are divided into 8 business lines (corporate finance, retail, asset management etc.) and these are charged separately in a fixed proportion (β) to their individual gross income. The proportion however differs across business lines to reflect the industry-wide averages of operational losses in particular areas of business activities**.**

Finally, the most sophisticated and most important approach is AMA - Advanced Measurement Approach - as it is the only approach allowing the banks to employ their own methodology of risk assessment. Of course that this freedom goes only as far as the benevolence of the supervisor – bank's methodology must strictly comply with the criteria set in the Basel II documentation. Generally the risk capital is evaluated for one year holding period at very high confidence level (currently set to 99.9%).

Basel Committee encourages banks to adopt AMA since it is the only approach reflecting their individual risk profiles (as opposed to aforementioned Standardized Approach and Basic Indicator Approach). The motivation of the banks to use AMA generally is that this approach rewards prudency and lower-than-average risk level by lower capital charge, hence creating competitive advantage. If they are able to do so banks are allowed to switch to more sophisticated approach but once they do there is no way back to the less sophisticated ones. Obviously it is expected that all the banks will gradually adopt AMA. Yet at the moment not all of the banks have the necessary qualifications for AMA mostly because of the operational risk management as such has been around for a relatively short time - for example a well-organized internal database of historical losses is needed (Basel II specifies a minimum of 5 years of a high-quality internal data) and indeed also some internal models of operational risk measurement and reporting must be established in order to switch to AMA.

Although original Basel II formulation gave a list of specific methods for obtaining the estimates of operational risk capital (see Chernobai et al. (2007) for overview), this was later left out and more recent documents have left the choice entirely to the banks' discretion. Nevertheless the set of requirements that the chosen methodology should fulfill is quite restrictive and technically most of the regulation and documentation – see e.g. Basel (2006) or BIS (2010a) - is directed to the method called Loss Distribution Approach (LDA) or its variants suggested in the original document. It is a method that evaluates operational loss and frequency distributions and extrapolates them in an actuarial type model to obtain estimates of operational risk exposure. This method is also a cornerstone of this paper and will be described in more detail later.

Basel II also does not single out any specific risk measures to be used under AMA – choice is at the discretion of the bank but it is subject to the approval of the supervisor. This state of affairs is a natural reflection of the fact that best practice for the operational risk measurement is still far from being recognized and established and that the discussion on the pertinence of individual methods is at its height. The most commonly used risk measure also recognized by Basel II is Value-at-Risk, but recent documentation also briefly mentions other risk measures – in BIS (2010, p.42) we can find the following:

> *"The most common and, so far, most adopted measure in risk management, including operational risk, is the Value at Risk (VaR). However, in certain applications and fields, including risk management, Shortfall measures (e.g. Expected Shortfall, Median Shortfall) have also gained notoriety and consensus due to their superior ability in representing the whole tail region and in providing a coherent risk estimate (under a sub-additivity perspective)."*

Let us now only remind that Expected Shortfall is the most important specimen of the class of "coherent risk measures", the one we are suggesting as a coherent alternative to VaR in our research. This is just another indication of growing awareness of this class of risk management tools as for example BIS (2001a, 2001b) documents do not mention any risk measures besides VaR.

Chapter 2. Modeling of operational risk

This chapter gives an in-depth description of the procedure for estimating the risk exposure employed in this paper. In a broader view it discloses the process leading from the recording the operational loss data to calculation of economic or regulatory capital charge using one or more chosen risk measures. Because of some of its specific features (such as focus on events of high magnitude and low frequency) operational risk modeling has a lot in common with insurance mathematics. It is hence quite natural that many concepts used in operational risk modeling originated in the field of insurance which has been rigorously researched for much longer time. This conjunction of quantitative risk management and insurance mathematics can be illustrated on many works that we are citing, e.g. Dowd and Blake (2006), Cizek et al. (2005), Wang (2002) , or the texts of professor Embrechts whose now notorious work on operational risk is rooted in his previous research of insurance mathematics (see e.g. Embrechts (1997)).

The specific method we are employing in this thesis is Loss Distribution Approach (LDA). Basics description of the methodology is given in the following section 2.2. Rest of the chapter then discusses the specific issues within LDA framework. But first we give some basic overview of other empirical research on operational risk to justify our model selection.

2.1. Literature overview

Due to scarcity and confidentiality of the data and also owing to the fact that operational risk is a very young discipline there is only little research working with real operational loss data. Indeed, some public databases of operational losses pooled across institutions are available, but these typically cover only short time periods and have a different nature than internal datasets of individual companies (yet typical empirical research on operational risk would employ such databases). Below we offer a brief overview of most relevant results with implications for LDA methodology.

Chernobai et al. (2005) tries to fit a large number of parametrical distributions to the operational loss data from a large public European database (they fit the distributions to several distinct sets of data divided by the event type[5]). They conclude that the data is characterized by heavy tails and that the thin-tailed distributions like Weibull or lognormal generally cannot provide a good fit. Heavy-tailed distributions like Pareto are evaluated as more suitable but none of them is able to consistently provide good results when fitted to different sets of losses, which highlights the idea that the choice of the optimal distribution is highly data-dependent.

Moscadeli (2004) analyzes a pooled dataset of operational losses from year 2001 of 89 different banks and tries to fit lognormal, Gumbel, Pareto, exponential and gamma distributions. Only lognormal and Gumbel provide a reasonable fit, nevertheless just for the body of distribution, not in the extreme quantiles. Moscadeli also employs Extreme Value Theory and finds that Generalized Pareto Distribution outperforms other distributional assumptions, hence making a strong point against the use of conventional statistical approaches.

Another example of successful use of Extreme Value Theory on operational risk data is the work of De Fontnouvelle et al. (2003) who fit a Generalized Pareto Distribution and also conclude that it outperforms all other fitted parametric distributions. De Fontnouvelle et al. (2005) further fits a wide list of distributions - exponential, Weibull, lognormal, gamma, gamma, Pareto, Burr, and loglogistic - to the operational loss data (for single banks sorted by event types) to confirm that it is characterized by fat tails and can only be modeled by heavy-tailed distributions. For example 1-parameter Pareto or gamma distribution has fitted the data for some loss event types quite well, but again the results have not been consistent and for many other sets of losses the models have been rejected by goodness of fit measures. They have also been able to reach better results with EVT modeling.

Also Rippel (2008) and Chalupka and Teply (2008) who analyze the same internal dataset of a commercial bank reach the conclusion that conventional distributional approach produces poor results. They however reach slightly different conclusion on the use of EVT – while Chalupka and Teply focus on the fit on the dataset and report great

[5] Overview of operational risk event types is given in Table 1.1 in the first chapter.

results, Rippel concentrates on the stability of results in scenario analyses and finds EVT overly sensitive to the extreme events, with a tendency to severely overestimate the regulatory and economic capital charges.

Some rather encouraging evidence has been gathered on the usefulness of g-and-h distribution for operational risk modeling. Dutta and Perry (2007), Jobst (2007) and Rippel (2008) find g-and-h very suitable distribution for the purposes of operational losses, with results superior to other parametric distributions[6] and surprisingly consistent even under various stress testing scenarios. On the other hand Chalupka and Teply (2008, p.23) report an unconvincing fit concluding that "whole distribution pattern of operational losses with rather limited observations is not possible to be captured even with a general class of distributions such as the g-and-h distribution."

Final remark concerns the typical features of the data. All studies above consistently report that operational risk data are characterized by significant leptokurtosis (i.e. heavy tails) and dominance of extreme events. Typically a couple of largest losses makes a substantial percentage of a total sum of all observations, it is not uncommon the largest losses might be tens of thousand times higher than the median of the data. That causes specific problems in operational risk models, especially since the risk measures are focused on extreme quantiles of loss distributions.

2.2. Loss Distribution Approach

Amongst the Advanced Measurement Approaches under Basel II the Loss Distribution Approach is by far the most important method of quantitative operational risk modeling – both amongst practitioners and academics (e.g. Chernobai et al. (2007)[7]). But in fact all of the operational risk empirical research that we cite in this thesis uses some form of LDA). LDA can be thought of as a comprehensive technique used to advance from the dataset of operational losses to the dollar value of risk exposure. LDA is used under AMA to calculate the regulatory capital charge but it can be used just as well for calculating the

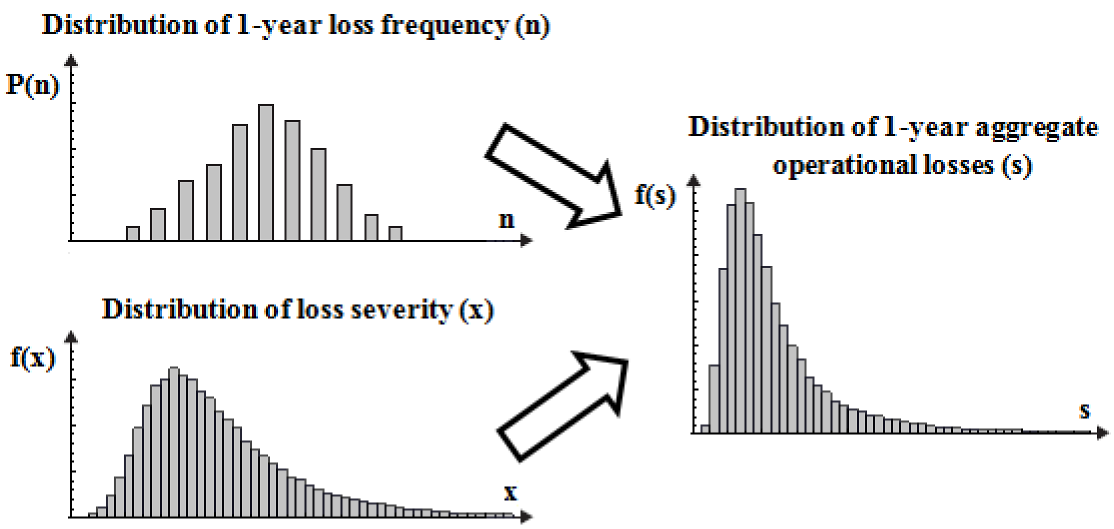
[6] That in case of Rippel (2007) includes also the extreme value theory distributions. For those author concludes that they fail on account of insufficient number of extreme observations. Jobst (2007) finds both g-and-h and EVT plausible suggesting that they could be used in complementary fashion to study the tail risk profile.

[7] Even though the loss distribution approach is a predominant one, some alternatives are possible. This reference provides an exhaustive overview with links to other literature.

economic capital. We will later argue that the only (but crucial) distinction between regulatory and economic capital estimation is the risk measure employed in the last step, i.e. in calculation of the risk exposure from the aggregate loss distribution.

Having the operational loss data one needs to make assumptions on their distribution in order to be able to proceed with modeling and estimations. It is customary in operational risk literature to consider two distinct aspects of the data – frequency and severity - and model them separately. Two estimated distributions are then joined (numerically using a Monte Carlo simulation engine) into a distribution of aggregate operational losses over the chosen time period (1 year), as schematized in Figure 2.1. Finally a relevant risk measure is applied on the aggregate loss distribution to obtain a dollar estimate of an operational risk exposure and corresponding capital charge. In the following subsections we are going through this process step by step in deeper detail.

Figure 2.1: Diagram of aggregate loss distribution estimation



Source: Author based on Chernobai et al. (2007).

2.3. Frequency distribution of operational losses

One of the specifics of the operational loss data is its time structure. Let us first consider market risk – it is conceptually straightforward to construct an empirical distribution of profits and losses of a given portfolio as we can easily observe its performance in almost any time interval we choose (even in real time). In case of operational risk the situation is very different and one needs to think not only about the distribution of magnitude of losses but also about their timing. One indeed cannot exactly predict the timing of operational losses they arrive in irregular intervals[8] and their frequencies vary a lot depending on the nature of the loss – while some small-scale losses like example minor credit card frauds might arrive frequently, say on daily basis, other losses are quite rare (e.g. property write-downs due to natural disaster) and might occur once in several years or even decades. It follows from this fact that the construction of loss distribution must be somewhat different compared to above-mentioned market risk as it should somehow reflect the specifics of time arrival of various operational losses.

Of course that modeling of the frequency presupposes a consistent methodology in the collection of the data within the bank. Specifically one is interested in meaning and consistency of dates of individual records in the database for banks can record them either as dates of events that caused the loss, dates of identification of the loss or accounting dates (see Chernobai et al. (2007) for details), in the worst-case scenario it might be some combination of the three. Underlying collection methodology should be examined closely because various methods might have different implications for modeling of frequency - for example in the case that dates are recorded as accounting dates, i.e. dates when the operational loss is realized and the event is assumed closed, one can expect clustering of the records around the ends of accounting periods, months or years.

Poisson distribution

Denote λ the mean number of events to arrive during any specified time period. Then the probability that k events will arrive within that time period can be expressed using Poisson distribution, which is defined in the following way:

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}; \quad k = 0,1,\ldots$$

$$with\, E(X) = \lambda\ and\, var(X) = \lambda.$$

All that is needed to model the frequency distribution as a Poisson process is to choose a time interval and estimate the λ parameter hence it is very easy to implement. Moreover Poisson distribution has the following interesting property:

[8] Yet typical internal dataset of operational losses seems to bear some regularities in the data frequency, like clustering of losses at the end of the year and months, possibly due to closing accounting periods.

Let X and Y be two independent Poisson random variables with parameters λ_X and λ_Y, then the distribution X+Y is a Poisson process with parameter $\lambda_X + \lambda_Y$.

That is very useful for the operational risk modeling because as long as we can assume independence of various business units of the bank we can model the frequency distribution of their losses as a single Poisson random variable.

Most of the empirical studies of operational risk manage with modeling the loss frequency using a simple Poisson or negative binomial distribution (which is basically a Poisson distribution where λ is a random variable of a gamma distribution instead of a constant) – for some examples see e.g. de Fontnouvelle et al. (2003) or Chernobai et al. (2007) where also a review of more sophisticated methods can be found. Even though some other more complicated approach could possibly provide a better fit on the empirical frequency of data we are also using Poisson distribution approach in our research as it still describes the data fairly well. Following subsection proceeds with the description of modeling of severity distribution.

2.4. Severity distribution of operational losses

A myriad of quantitative techniques has been developed for modeling of financial data. In this section we are presenting some of the methods that are common in operational risk modeling and that we are using in the empirical part of this thesis. There are two main ways to tackle the problem of modeling the severity distribution of operational losses – non-parametric approach and parametric approach[9].

2.4.1. Non-parametric approach

Central idea of this approach is to base the modeling of the losses on their empirical distribution while trying to avoid any additional assumptions if possible. A natural estimate of the loss distribution function is the observed distribution of loss size. Cumulative empirical distribution function can be expressed this way:

$$P(X \leq x) = \frac{\sum_{j=1}^{n} \boldsymbol{I}_{X_j \leq x}(X_j)}{n},$$

[9] In the terminology we are following Chernobai et al. (2007) but some other authors, for example Burnecki et al. (2010), prefer the terms "empirical" and "analytical".

$$\text{with } I_{X_j \le x}(X_j) = \begin{cases} 1 \; if \; X_j \le x \\ 0 \; if \; X_j > x \end{cases},$$

where *n* is the number of observations. It is a step-wise function with jumps of size[10] 1/*n* occurring at each observed amount X_j. Burnecki et al. (2010) append that, especially for the large samples, the empirical distribution function is often approximated by a piecewise linear function with the "jump-points" connected by linear functions.

Empirical distribution is only appropriate when there is a sufficiently large database of losses available and its use requires two assumptions (Chernobai et al. (2007)):

1) Historic loss data are sufficiently comprehensive.
2) Past losses are equally likely to reappear in the future, and losses of other magnitudes (such as potential extreme events that are not a part of existent database) cannot occur.

Considering the idiosyncraticity of operational risk events such conditions are extremely unlikely, especially for the tails of distributions. Exceptionally large losses are possible and the data on those is nearly always scarce and incomprehensive – moreover it would be very dubious to assume that such events are going to reoccur with the same probability and same severity.

Even though most of the times this approach is perceived as insufficient and inferior to parametric approaches, we should note that the concept of empirical distribution is still very useful as it is often used to check the goodness of fit of estimated parametrical distribution. Measuring the distance between the empirical and fitted analytical distribution function is one of the most natural tests for goodness of fit – if the estimated distribution follows the empirical distribution closely it indicates a good fit and vice-versa (see Chernobai et al. (2007) or Burnecki et al., 2010). Examples of such measures of goodness of fit are e.g. Kolmogorov-Smirnov (KS) test or Anderson-Darling (AD) test (they are introduced in section 2.4.4).

2.4.2. Parametric approach

Contrary to the previous method this one relies on ex-ante assumptions on the functional form of the loss distribution function. After choosing some theoretical

[10] Indeed this holds only when $X_i \neq X_j$ for $\forall i, j: i \neq j$.

distribution one evaluates the values of its parameters using the available data to obtain best possible fit. It should be stressed right on the beginning that due to the nature of the operational loss data many standard models commonly used in statistics are not appropriate for fitting the severity distributions. First, due to the definition of the operational risk given earlier only the distributions with positive domain of definition are admissible. Second, due to the fact that operational risk data is typically driven by low-frequency high-severity losses any light-tailed distributions (i.e. distributions with fast decay of probability density in the tails, such as normal distribution) are typically rendered useless. This conclusion is confirmed also in the empirical part of this thesis.

Even though there is no consensus over the “best” severity distribution (and there probably cannot be as the most appropriate distributional form is data-dependent) most of the literature coincide in the conclusion that critical task in operational risk modeling is to properly capture the behavior of the tails of loss severity distribution (by tails we mean the very high quantiles, i.e. those parts of distribution that describe the probability and of extreme events). This is reinforced by the fact that to assess operational risk exposure we are using risk measures that employ extremely high quantiles of the distribution of losses (such as 99.9% value-at-risk for regulatory capital charge under Basel II) hence we need to model the tail of the distribution as precisely as possible.

Some conventional distributions that are most common in loss severity modeling are exponential, lognormal, Weibull, Pareto, beta and gamma distributions or more recently a four-parameter g-and-h distribution (which we include as an example of the generalized parametric distributions that are sometimes used). In empirical research authors would typically try to fit several distributions and then choose the best using various goodness of fit measures (see section 2.4.4). We adopt the same approach in the empirical part.

Owing to the definition of operational risk, only the distributions defined for $x>0$ can be used for operational risk modeling. Since some of these distributions are not well-known we include in the following section a short review of the ones we use in the empirical part of the thesis.

Exponential Distribution

Exponential distribution is one of the simplest parametrical distributions used in quantitative risk management, e.g. Burnecki et al. (2010) state that it is often used in developing models of insurance risks due to its useful mathematical properties. Its probability density and cumulative distribution functions contain a single parameter $\lambda>0$ and are defined in the following way, respectively:

$$f(x) = \lambda e^{-\lambda x}, \;\; x > 0,$$
$$F(x) = 1 - e^{-\lambda x}, \;\; x > 0.$$

The maximum likelihood estimator (MLE) for λ is a simple empirical mean of the data, hence:

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Potential problem in the context of operational risk modeling is the exponential decay in the right tail which makes this distribution implausible for modeling of heavy-tailed data since the high-severity losses are given a near-zero probability (Chernobai et al. (2007). Implementation in Monte Carlo simulation engine is trivial since random variates can be generated using the inverse transform method by $X = -\frac{1}{\lambda}\log U$ where U is a random variable distributed uniformly on the $(0,1)$ interval[11].

Lognormal Distribution

Consider the transformation of standard normal random variable $\Phi(X)$ with density and distribution functions given by:

$$f(x) = \frac{1}{\sigma x\sqrt{2\pi}} e^{-\frac{(log\, x-\mu)}{2\sigma^2}},$$
$$F(X) = \Phi\left(\frac{log\, x - \mu}{\sigma}\right), \;\; x > 0.$$

This distribution is known as lognormal. The parameters $\mu\epsilon\mathbb{R}$ and $\sigma > 0$ are the location and scale parameters respectively. Their maximum likelihood estimators can be defined as follows:

[11] This useful fact stems from the fact that the inverse of the distribution has the form of $\mathrm{F}^{-1}(\mathrm{p}) = -\frac{1}{\lambda}\log(1-\mathrm{p})$ with $\mathrm{p} \in (0,1)$.

$$\hat{\mu} = \frac{1}{n}\sum\nolimits_{i=1}^{n} log\, x_i\,, \quad \hat{\sigma} = \sqrt{\frac{1}{n}\sum\nolimits_{i=1}^{n} (log\, x_i - \hat{\mu})}.$$

Equivalently, one can obtain same parameter estimates by fitting normal distribution to the natural logarithms of original data. Lognormal distribution has been suggested for operational risk modeling e.g. in BIS (2001b), but it is also a thin-tailed distribution even though the decay in the right tail is slower than for exponential distribution with[12] $\bar{F}(X) \sim x^{-1} e^{-log^2 x}$. In most empirical applications it was not implemented with a great success because it is not very good for modeling the heavy tails as a typical feature of operational risk data. Finally the random variates[13] from lognormal distribution can be simulated as $X = e^{\Phi^{-1}(U)\sigma+\mu}$, where Φ is the standard normal distribution (Chernobai et al. (2007)).

Weibull Distribution

A probability law with density and distribution functions defined as:

$$f(x) = \tau\beta x^{\tau-1} e^{-\beta x^{\tau}}, \;\; x > 0,$$
$$F(X) = 1 - e^{-\beta x^{\tau}}, \;\; x > 0,$$

with τ >0 being a shape parameter and $\beta > 0$ a scale parameter, is known as Weibull (or sometimes Frechet) distribution. It is a generalization of exponential distribution – if random variable V is exponentially distributed, then its transformation $X = V^{1/\tau}$ is from Weibull distribution (obviously Weibull reduces to exponential distribution for $\tau = 1$). According to Chernobai et al. (2007) Weibull distribution has been found to be the optimal distribution in reinsurance models as well as in asset returns models. New parameter adds greater flexibility and allows for heavier scales - for values of τ smaller than 3.6 the distribution is right-skewed[14] and the decay in the right tail is slower (for $\tau < 1$ Weibull distribution is actually heavy-tailed).

Random variates from Weibull distribution can be obtained by the transform method – first generate the random variates Z from exponential distribution using parameter estimate $\hat{\beta}$ and then take the transformation $X = Z^{1/\hat{\tau}}$ to obtain Weibull random variables.

[12] $\bar{F}(x)$ refers to a survival function, defined as 1-F(x).
[13] It is again derived from the inverse distribution function$\mathrm{F}^{-1}(\mathrm{p}) = \mathrm{e}^{\Phi^{-1}(\mathrm{p})\sigma+\mu}$, where p is probability.
[14] For $\tau \approx 3.6$ the distribution is approximately symmetrical.

Gamma Distribution

A random variable is said to be from gamma distribution if its density and distribution functions are as follows[15]:

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

$$F(X) = \Gamma(\alpha;\; \beta x), \;\; x > 0.$$

Parameters $\alpha > 0$ and $\beta > 0$ are shape and scale parameters respectively. According to Burnecki et al. (2010) the gamma law is one of the most important distributions for modeling in insurance finance. Useful mathematical properties and details on estimation and generating random variates from gamma distribution (which is not very straightforward) can be found in Cizek et al. (2005).

Beta Distribution

The following density and distribution functions specify the beta probability law[16]:

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 \le x \le 1,$$

$$F(X) = I(x;\; \alpha, \beta), \;\; 0 \le x \le 1.$$

Parameters $\alpha > 0$ and $\beta > 0$ determine the shape of the beta distribution. First thing to notice about this distribution is that its support is bounded by the closed interval $[0,1]$. Hence in order to model operational risk data with beta distribution one needs to first rescale them to fit this interval. For a known (chosen) data-scaling parameter $\theta > 0$, such that $x_i \le 1, i = 1, .., n$ the beta distribution can take the following form:

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

$$F(X) = I\left(\frac{x}{\theta};\; \alpha, \beta\right), \;\; 0 \le x \le \theta.$$

[15] Integral $\Gamma(\alpha)$ is called a complete gamma function and is defined as $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$. Integral $\Gamma(\alpha;\; \beta x) = \frac{1}{\Gamma(\alpha)} \int_0^{\beta x} t^{\alpha-1} e^{-t} dt$ is known as incomplete gamma function.

[16] Term $I(x;\; \alpha, \beta)$ in the definition of the distribution function is the regularized beta function defined as $I(x;\; \alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \int_0^x u^{\alpha-1}(1-u)^{\beta-1} du$.

Details on estimation and random number generation can be found in Chernobai et al. (2007).

Pareto Distribution

Pareto distribution is defined by the following density and distribution function[17]:

$$f(x) = \frac{\alpha\beta^{\alpha}}{(\beta + x)^{\alpha+1}}, \quad x > 0,$$

$$F(X) = 1 - \left(\frac{\beta}{\beta + x}\right)^{\alpha}, \ x > 0.$$

Both scale parameter β and shape parameter α are positive. Pareto distribution has a particularly thick tail (the lower the α the thicker the tail), which is in fact monotonically decreasing as can be seen from the survival function: $\bar{F}(x) = \left(\frac{\beta}{\beta+x}\right)^{\alpha}$ (because tails of Pareto distribution are governed by the power function they are sometimes called "power tails", as opposed to exponentially decaying tails). When $\alpha \leq 1$ Pareto distribution has infinite mean and variance, hence the tail is extremely fat and losses of infinite magnitude are possible. If such distribution is used for the generation of aggregate losses, no quantile-based risk measure such as VaR or Expected Shortfall can tell us anything useful about the risk exposure as the outcomes are somewhat random. This property is a potential problem, otherwise Pareto distribution seems as a promising tool for operational risk modeling for its ability to account for heavy tails.

G-and-h distribution

In past few years several studies have generated good results modeling the operational risk data by four-parameter g-and-h distribution (these include Dutta-and-Perry (2007) or Rippel and Teply (2008), for a more theoretical treatment with simulated data analysis see Degen et al. (2007)). The g-and-h distribution was first introduced by Tukey (1977) and it is another transformation of standard normal variable Z defined as:

$$X_{g,h}(Z) = A + \frac{B}{g}(e^{gZ} - 1)e^{\frac{hZ^2}{2}}, \quad Z \sim N(0,1)$$

[17] Note that this is only one of the possible specifications of Pareto distribution. Generalizations with more parameters (but also a simplification with single parameter) are possible. Pareto distribution can also be reformulated into the Generalized Pareto Distribution used in EVT, as we will show later.

There are two advantages this distribution has to offer. Its main strength lies in its flexibility - with different combinations of values of the parameters (A, B, g, h>0) it can approximate a wide variety of distributions. Besides some of the distributions described above it can also probabilistically approximate the shape of the Generalized Pareto Distribution and Generalized Extreme Value distribution from Extreme Value Theory (see Hoaglin (1985) or Degen et al. (2007) for details). Secondly since it is a mere transformation of the standard normal distribution it is quite easy to generate random numbers from g-and-h for the purposes of Monte Carlo simulation.

A slight technical problem is the estimation of parameters since g-and-h distribution is typically not included in statistical packages. I have done the estimation using the percentile method algorithm from Hoaglin (1985). Assume that X is the vector of operational risk data, denote X_p the p-th percentile of the g-distribution and Z_p p-th percentile of standard normal distribution, then

$$\hat{A} = median(X) \ \ and \ \hat{g} = median\left[-\left(\frac{1}{Z_p}\right) log\left(\frac{X_{1-p}-X_{0.5}}{X_{0.5}-X_p}\right)\right].$$

Parameter estimates $\hat{B}$ and $\hat{h}$ are obtained from the following OLS regression:

$$log\left(\frac{g\left(X_{1-p}-X_{0.5}\right)}{e^{-gZ_p}-1}\right) = log(b) + h * \frac{Z_p^2}{2},$$

where $\hat{h}$ is the estimated coefficient and $\hat{B} = \exp(\hat{b}\ )$.

There is one more complication in the usage g-and-h distribution for data modeling - because the probability density functions and cumulative distribution functions for g-and-h distribution are generally unknown (Degen et al. (2007)) it is not possible to calculate the test statistics for Kolmogorov-Smirnov and Anderson-Darling test that we use for all other distributions as measures of goodness of fit. In this case we will need to rely on visual inspection of QQ plots and plausibility of calculated capital charges as our sole tools for model evaluation (see section 2.5.5 for details).

2.4.3. Extreme Value Theory Approach

Alternative to a classical distribution fitting are approaches derived from extreme value theory (EVT) – a statistical method developed to analyze data samples characterized by extreme events. We give only a basic summary of the method, for more detailed theoretical treatment reader might consult Embrechts et al. (1997). Some practical issues with EVT in operational risk modeling are described in Embrechts et al. (2003).

Basic idea is to model the large losses separately from the medium/small losses (that can be modeled in a traditional way) which allows us to describe the tails of the distribution more precisely. Indeed on the first glance this approach seems quite suitable for operational risk modeling because the extreme events are of the crucial importance. Also in our analysis EVT is employed because the traditional methods prove as insufficient to model the operational risk data.

Regardless of the chosen EVT method, body of the distribution can be modeled using one of the severity distributions described in previous section using all observations that were not selected as "extreme". Another common approach is to model only the tail using EVT while taking simply the empirical distribution function for the body of the losses. A slight complication hence arises for the simulation of the aggregate losses which must somehow take in account the two distinct distributions of severity – one for the small/medium losses and one for the large losses. Yet modeling the tail independently can be a huge help in operational risk modeling, because, as we demonstrate in Chapter 4, traditional distributions might have a very good fit in the body of the distribution of losses but they typically exhibit an extremely poor fit at the tail.

There are two basic approaches to EVT that differ in a way they distinguish between large and small-to-medium losses, i.e. between those data that will be modeled using EVT and those that will not. The two approaches are called Block Maxima Method and Peaks over Threshold Method.

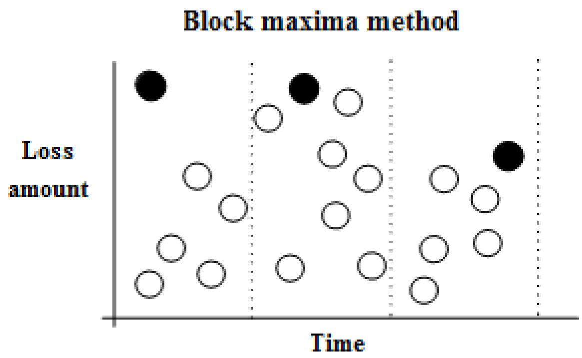
Block Maxima Method

Block Maxima Method (BMM) divides the sample into subsamples of equally long time periods (blocks) and selects the highest value observation from each block. Selected observations then constitute the extreme events. According to Chernobai at al. (2007) the limiting distribution of these observations is the generalized extreme value distribution (GEV) whose probability density distribution function can be written as follows:

$$f_{(\mu,\sigma,\xi)}(x) = \left(\frac{\xi(x-\mu)}{\sigma} + 1\right)^{-\frac{\xi+1}{\xi}} e^{-\left(\frac{\xi(x-\mu)}{\sigma}+1\right)^{-1/\xi}} \quad for\ \xi(x-\mu)/\sigma+1>0,$$

and zero otherwise, where x are the block maxima observations and the three parameters μ, σ, ξ denote location (any real number, but often assumed to be equal to zero), scale (any positive real number) and shape parameter respectively. Shape parameter governs the tails of the distribution and, contrary to shape parameter α in Pareto distribution, larger ξ indicates heavier tails (Moscadelli (2004)).

Figure 2.2: Selection of extremes in Block Maxima Method



Source: Author.

Peaks over Threshold Method

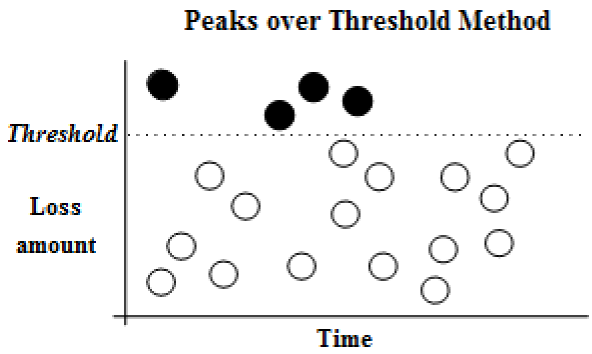
Peaks Over Threshold Method (POTM) selects the extreme events as observations exceeding certain high threshold. For x being the observations exceeding the threshold the limiting distribution is Generalized Pareto Distribution (Chernobai et al. (2007)) with probability density function of

$$f_{(\mu,\sigma,\xi)}(x) = \frac{1}{\sigma}\left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{\left(-\frac{1}{\xi}-1\right)},$$

where, as in the case of BMM, the parameters μ, σ, ξ are those of location, scale and shape respectively. Also the interpretation of shape parameter is the same. Note that $\xi > 1$ indicates infinite mean model which causes same troubles for the risk exposure estimation as in the case of Pareto distribution, see section 2.5.3 for details.

The crucial issue for POTM is of course the selection of the appropriate threshold (we shall denote it by u). Unfortunately there is no widely accepted analytic solution for doing so and often the threshold is arbitrarily chosen to cut-off certain percentage of the highest losses, e.g. 5% or 10% largest losses. Although there is no guaranteed way to select the “correct” threshold the estimated parameter values are known to be sensitive to the threshold choice. Below we describe one purely visual method that might help with the threshold selection (unfortunately it does not prove very useful in empirical part).

Figure 2.3: Selection of extremes under Peaks Over Threshold Method



Source: Author.

Threshold selection using mean excess plot

One way to select the high threshold for isolating the extreme losses is to resort to a visual inspection of a so-called “mean excess plot”. Mean excess function is defined as:

$$e(u) = E(X - u|X > u),$$

i.e. for a threshold u it gives a mean value of differences between u and losses X for X exceeding u. Mean excess plot then displays values of mean excess function as a function of u. If the data follow GPD, the plot should be linear in u as its mean excess function clearly shows:

$$e_{GPD}(u) = \frac{\beta + \xi u}{1 - \xi}, \qquad (\beta + \xi u) > 0,$$

hence if we select such threshold u, above which the plot is roughly linear, to find a candidate set of extreme losses that might be expected to follow GPD.

2.4.4. Goodness of fit tests

Critical task in operational risk modeling is to select a correct model. In loss distribution approach that generally means that we need to find a statistical distribution that can model the data reasonably well. Mistakes can be costly as the chosen model directly determines the capital charge and can result in serious overestimation or underestimation of the risk exposure. In order to reduce this model risk we can use various goodness of fit tests (in ideal case a combination of them).

Perhaps the most intuitive way to test the fit of the distribution function to the data is to compare the empirical and theoretical distribution function. Visually it can be done using the QQ plot with empirical quantiles of the data on horizontal axis and theoretical quantiles of the fitted distribution on vertical axis. In case of perfect fit the quantiles should be equal, i.e. the points in the plot should follow a 45 degree line. We will also review and later employ two formal tests[18] based on empirical distribution function - Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) tests.

Denote $F_n(x)$ the empirical distribution function and $F(x;\,\hat{\theta})$ the fitted distribution function where $\hat{\theta}$ is the vector of estimated parameters. Both formal tests described below test the same null hypothesis against the same alternative:

$$H_0\colon F_n(x) = F(x;\,\hat{\theta}),$$
$$H_1\colon F_n(x) \neq F(x;\,\hat{\theta}),$$

i.e. under the null the data do follow the specified parametric distribution function. In these tests, low values of test statistic T represent the evidence in favor of the null hypothesis whereas large ones indicate its invalidity. We report the p-value defined as:

$$p - value = P(T \geq t),$$

where t is the critical test value. It tells us the probability of obtaining T at least as large as observed outcome if the null was true. Hence for small p-values we reject the null and for large ones we do not reject it. It is important to note that in case of Anderson-Darling test the critical test values might change depending on the tested parametric distribution.

Kolmogorov-Smirnov test

KS test finds the largest vertical distances between the empirical and the fitted distribution functions and use them to calculate the test statistics, which is then compared to critical values of Kolmogorov distribution. Formally:

[18] There are also other families of tests used to assess goodness of fit, such as Chi-squared tests, but the KS and AD are by far the most common amongst tests used in operational risk modeling.

$$D^{+} = \sup_{x} \{F_n(x) - F(x;\ \hat{\theta})\},$$

$$D^{-} = \sup_{x} \{F(x;\ \hat{\theta}) - F_n(x)\},$$

$$KS = \sqrt{n}\, max\{D^{+}, D^{-}\}.$$

For direct computation of KS statistic following computing formula can be used[19]:

$$KS = \sqrt{n}\, max\left\{\sup_{j}\left\{\frac{j}{n} - z_{(j)}\right\}, \sup_{j}\left\{z_{(j)} - \frac{j-1}{n}\right\}\right\},$$

$$z_{(j)} = F(x_{(j)})\ for\ j = 1,2,\dots,n; x_{(1)} < x_{(2)} < \cdots < x_{(n)}.$$

From the definition of KS it is clear that it puts the same weight to the fit of each quantile of the distribution, hence it is it is a somewhat crude tool for evaluation of overall fit of the distribution function. A good complementary test to KS is Anderson Darling test which we are also employing in the empirical part.

Anderson-Darling test

We are using the so-called quadratic type AD test, where the test statistic is defined as follows:

$$AD^2 = n\int_{-\infty}^{\infty} \frac{\left(F_n(x) - F(x)\right)^2}{F(x)\left(1 - F(x)\right)} dF(x)$$

The computing formula for quadratic AD test can be defined as:

$$AD^2 = -n + \frac{1}{n}\sum_{j=1}^{n}(1-2j)\, log\, z_{(j)} - \frac{1}{n}\sum_{j=1}^{n}\left(1 + 2(n-j)\right) log\left(1 - z_{(j)}\right).$$

According to Chernobai et al. (2007), AD test tends to put most weight on the tails of the distribution, unlike KS test. Thus the AD test proves very useful when the underlying data are heavy-tailed and one wants to put emphasis on the fit in the extreme right quantiles, which is typically a case in operational risk data modeling.

[19] It can be found using probability integral transformation method, see Chernobai et al. (2007) for more detailed theory.

2.5. Aggregate loss distribution

Granted that we were able to estimate frequency and severity of the operational losses we still need to aggregate them into a single joint loss distribution; one common way is to use actuarial models (see Chernobai pp.223-225 for details and description of other methods). Several assumptions are necessary when applying this procedure. Let *n* be the number of loss events and $\{X_1, X_2, \ldots\}$ the realized loss amounts.

(i) Conditional on *n*, X_i are independent identically distributed positive random variables.

(ii) The distribution of X_i is independent from *n*.

(iii) The distribution of *n* is independent from X_i.

The prevalent method for numerical evaluation of compound distribution of losses which is also used in this paper is relying on Monte Carlo simulations. It follows these steps:

1) Choose a time interval Δt and number of simulations (scenarios) N. For each scenario draw a random number from the underlying fitted loss frequency distribution *g* to simulate the number of losses and get:

$$n_{i,\Delta t}, \qquad i = 1, \ldots, N, \qquad n_{i,\Delta t} \sim g_{\Delta t}(n)$$

2) For each scenario simulate $n_{i,\Delta t}$ loss amounts *s* using the loss severity distribution *f*, you get

$$\{s_{i,1}, s_{i,2}, \ldots, s_{i,n_{i,\Delta t}}\}_i, \quad i = 1, \ldots, N, \quad s_{i,j} \sim f(s)$$

3) Aggregate the losses for each scenario by summing them up to obtain N simulated total aggregated losses over the chosen time period:

$$S_{i,\Delta t} = \sum_{j=1}^{n_{i,\Delta t}} s_{i,j} \ , \quad i = 1, \ldots, N$$

4) Sort the values $S_{i,\Delta t}$ in increasing order to obtain the desired aggregate loss distribution.

Notice that (perhaps with exception of some special cases) such distribution function typically cannot be expressed analytically – it generally has no closed form as it involves an infinite sum over all possible values of $n_{i,\Delta t}$ where j-th term is weighed by a probability $P(n = n_{i,\Delta t})$ and involves n_i-fold convolution of the chosen severity distribution conditional on $n = n_{i,\Delta t}$ (see Peters and Sisson, 2006, for details). Different distributions can also be more or less demanding on the number of simulations, but generally the precision of the estimates clearly increases as N is getting higher (especially regarding the heavy tails because many observations are needed to generate enough extreme events). Chernobai (2007) suggests 100 000 repetitions, BIS (2010a, p.41) vaguely states that "the number should be consistent with the shape of the distributions and with the confidence level to be achieved."

With the aggregate loss distribution at hand we can finally apply a chosen risk measure to evaluate the risk exposure, for example Value-at-Risk is obtained simply as empirical α-percentile of the distribution where α is a desired level of confidence. Because the risk measures are discussed extensively in Chapter 3 it is not necessary to get into a further detail at this point.

We would only like to point here to the clear fact that the focus of operational risk modeling lies on the extreme events. We have noted earlier that operational risk data are driven by losses of huge severity and low frequency that dwarf other (expected) losses and represent the biggest threat for the banks. It is precisely these occasional losses that need to be modeled as accurately as possible so that the bank's capital reserve is sufficiently high to avoid insolvency should such a loss occur. The Basel II logic of having a capital buffer that covers 99.9% of 1-year aggregate loss is driven by the same considerations. Though concerning such extreme quantiles, only when we have a model that provides a convincing fit for the tails of the loss data distribution we can draw useful conclusions from the derived aggregate loss distribution.

Chapter 3. Risk measures and coherence

This chapter is devoted to the theoretical description of the risk measures we are using in empirical part. We first present a theoretical framework within which we define the concept of coherence. We than refer to some properties of Value-at-Risk and compare them with Expected Shortfall as a representation of coherent risk measures. That allows us to draw some conclusions on the use of VaR and ES in the context of operational risk modeling.

3.1. Concept of coherence

Let us now embed our considerations on the operational risk capital reserves calculation into a more rigorous framework. We generally follow the notation and logic of Artzner et al. (1999) seminal paper but modify its theoretical setup slightly to fit the context of operational risk (some adjustments were necessary mainly due to the fact that Artzner et al. consider two-sided risk definition with both positive and negative outcomes while in operational risk only losses are considered).

Consider a probability space $(\Omega, \mathcal{F}, \mathcal{P})$. Ω is interpreted as the set of states of nature at the end of a fixed reference time period (over which we want to quantify the risk hence in the typical operational risk exercise it would be 1 year) and is assumed to be finite. Denote $\bar{\mathcal{G}}$ the set of positive-real-valued random variables defined on the given probability space and assume:

(i) $X_1, X_2 \in \bar{\mathcal{G}} \Rightarrow X_1 + X_2 \in \bar{\mathcal{G}}$

(ii) $X \in \bar{\mathcal{G}} \Rightarrow aX \in \bar{\mathcal{G}}\ for\ all\ a > 0$

Further denote $\mathcal{G}$ an extension of the set $\bar{\mathcal{G}}$ such that $\bar{\mathcal{G}} \in \mathcal{G}$ and

(iii) $X \in \bar{\mathcal{G}} \Rightarrow X + b \in \mathcal{G}\ for\ all\ b \in \mathbb{R}$

Definition 1. **Risk measure**. *A mapping $\rho: \mathcal{G} \rightarrow \mathbb{R}$ is called a risk measure.*

A mapping $\rho: \mathcal{G} \rightarrow \mathbb{R}$ which assigns a real value to every element of $\mathcal{G}$ for any realized state of nature is called a risk measure. $\mathcal{G}$ is an extended set of all operational risks. Extension (iii) is needed to make sure that also all risks adjusted for some amount of capital charge b are in the domain of definition of ρ. Elements of $\mathcal{G}$ can be interpreted as aggregate operational losses at the end of the reference period. They are random variables representing the "final net worth[20]" of the bank's operational risk events under any possible state of the world.

The idea of Artzner is to define the set of acceptable final net worths and to measure the risk as a kind of "distance" of our position from an acceptable state - it is the amount that when added to the bank's assets ensures that its position is acceptable to the supervisor (be it manager, shareholder or regulator). Existence of certain reference instrument allowing for risk-free investment is also assumed so there always *is* an opportunity to add "prudence assets" safely. In the further inquiry we call the set of final net worths accepted by the regulator or supervisor "acceptance set" end denote it by $\mathcal{A}$. Following Artzner et al. let us now define the correspondence between acceptance sets and risk measures:

Definition 2. Risk measure associated with an acceptance set. The risk measure associated with the acceptance $\mathcal{A}$ is the mapping from $\mathcal{G}$ to $\mathbb{R}$ denoted by $\rho_{\mathcal{A}}$ and defined by

$$\rho_{\mathcal{A}}(X) = \inf\{m | m + X \in \mathcal{A}\}.$$

Definition 3. Acceptance set associated with a risk measure $\boldsymbol{\rho}$. The acceptance set associated with a risk measure ρ is the set denoted by $\mathcal{A}_{\rho}$ and defined by

$$\mathcal{A}_{\rho} = \{X \in \mathcal{G} |\ \rho(X) \leq 0\}.$$

These two definitions complete the basic framework set in Artzner et al. (1999). It should be clear now that higher ρ is associated with riskier positions. First definition says nothing else that given some acceptance set $\mathcal{A}$ risk measure is the smallest cash (capital

[20] But to avoid confusion we need to clarify that "worth" is defined here as a loss.

reserve) amount that makes the position acceptable. Second definition says that such positions are acceptable for which the risk measure returns a non-positive value.

If X is high enough the bank might face the insolvency. To protect itself from insolvency bank creates economic capital reserves (or the regulator imposes the minimum regulatory capital requirement) determined by the risk measure $\rho(X)$ which denotes the minimum level of available capital in the bank made ready to cover potential losses (such that the overall position of the bank – including capital buffer – is acceptable from the point of view of bank policyholder or the regulator). Essentially the $\rho(X)$ should be chosen such that probability of occurrence of event $X > \rho(X)$ is very small. Negative values of $\rho(X)$ can then be interpreted as amount of capital that can be reallocated from the operational risk capital reserve while preserving the acceptability of the position.

Following section introduces and discusses some properties of risk measures defined on $\mathcal{G}$. These were proposed by Artzner et al. (1999) in the form of axioms that – when satisfied – form the concept of coherence. Axioms are labeled with letters for further reference.

Definition 4. Coherence. *Let* $\mathcal{G} = \{X_1, X_2, X_3 \dots\}$ *be the risk set and* $\rho: \mathcal{G} \to \mathbb{R}$ *be the risk measure. The risk measure is coherent if it satisfies the following four axioms:*

1) **Axiom T. Translation invariance**. *For all* $\alpha \epsilon \mathbb{R}$ *and all* $X \epsilon \mathcal{G}$ *we have*
$\rho(X - \alpha) = \rho(X) - \alpha$.
2) **Axiom S. Subadditivity**. *For all* $X_1, X_2 \epsilon \mathcal{G}$ *we have*
$\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$.
3) **Axiom PH. Positive homogeneity**. *For all* $\alpha \geq 0$ *and* $X \epsilon \mathcal{G}$ *we have*
$\rho(\alpha X) = \alpha\, \rho(X)$.
4) **Axiom M. Monotonicity**. *For all* $X_1, X_2 \epsilon \mathcal{G}$ *such that* $X_1 \leq X_2$ *we have*
$\rho(X_2) \leq \rho(X_1)$.

Let us now briefly discuss the axioms. Axiom T simply states that subtracting a sure initial amount (i.e. decreasing risk by that amount) decreases the risk measure by the same amount and vice versa. When we set aside α dollars (or create capital reserve of α) to cover the potential losses the measure of our risk must decrease by α. This axiom flows naturally from the definition of risk measure and it has several implications. First it ensures that $X, \rho(X)$ and α must be all in the same units (currency units). It also implies that $\rho\big(X - \rho(X)\big) = \rho(X) - \rho(X) = 0$, which means that risk is fully offset (position is made acceptable) by investing sure initial amount equal to the measured magnitude of risk. That is important part of the definition which reflects the supervisor's point of view as for example the regulatory capital charges work exactly this way – we compute the magnitude of risk and require banks to create reserves in the same amount assuming that such precaution mitigates the impact of the risk.

"Merger does not create extra risk," says Artzner et al. (1999, p.209) regarding Axiom S. It is a natural requirement for a risk measure that it should comply with the diversification principle. This axiom is really the heart of the whole concept of coherence. In fact most of the risk measures that fail the test for coherence fail precisely at this point (VaR is one and the most important example). Implications of non-compliance with axiom S will be discussed throughout the paper, for now let us only summarize authors' rationale for its inclusion, apart from capturing the diversification principle. If a risk measure is not subadditive, then following paradoxes (from the point of view of supervisor or regulator) might occur:

- banks unable to meet the capital requirement implied by the risk measure might be motivated to split up into separate affiliates
- banks might be motivated to prefer less diversified (or even undiversified) portfolios and investment strategies[21]
- suppose that two business lines of single bank separately evaluate their risk $\rho(X_1)$ and $\rho(X_2)$. If ρ is subadditive than $\rho(X_1) + \rho(X_2)$ is an upper bound of the total risk $\rho(X_1 + X_2)$. Supervisor can be certain that imposing capital requirement of $\rho(X_1) + \rho(X_2)$ is sufficient without having to evaluate the total risk of a bank as a whole, that is $\rho(X_1 + X_2)$. Sum of the risks is representative of the total risk, unlike in case of non-additive measures where it says nothing at all.

For $n = \{1,2 \dots\}$ Axiom S implies that $\rho(nX) \le n\rho(X)$. Axiom PH describes the limit case of subadditivity where the diversification effect is exactly zero, i.e. where $\rho(nX) = n\rho(X)$.

Regarding Axiom M note that $X_1 \le X_2$ means that X_2 is a strictly better performing portfolio (with higher net worth in any realized state of the world). This axiom states that

[21] This point mainly concerns market risk though.

the better (stochastically dominant) portfolio cannot be associated with larger risk. It rules out certain risk measures based on standard deviation alone, where volatility determines riskiness with no regard to expected performance of the portfolio.

A fifth axiom of Relevance is sometimes stated: *for all risks X such that $X \geq 0$ we have* $\rho(X) \geq 0$. It simply says that if there is a certain loss it must be reflected in the risk measure. It is not included in a definition of coherence, but it is a "clearly necessary, but not sufficient, to prevent concentration of risk to remain undetected" (Artzner et al. 1999, p.210).

Notice that this axiomatic approach inherently defines risk measures as single dollar (or other currency) numbers. It is certainly virtuous in terms of simplicity and comprehensibility; on the other hand some authors are warning that describing risk by a single number can involve a great loss of information. Artzner et al. (1999) were well aware of this and claim on this account that even though some loss of information might be involved, there is no other way since "actual decision about taking a risk or allowing one to take it is fundamentally binary, of the yes or no type…" and "…this is the actual origin of risk measurement." Above argument can be however somewhat weaker in case of operational risk modeling since in this context we are typically not making a decision between taking and not taking the risk – what we are after is the evaluation of the risks we have already taken. But having a single dollar number as a risk measure still makes sense having in mind the aims of the regulator or internal supervisor - that is to make the bank's risk position acceptable by having enough reserve to cover for the potential operational losses, i.e. to setup the proper capital charge which by definition a single number. We will also later consider having a convenient *combination* of risk measures to keep as much information on underlying loss distribution as possible.

3.2. Value-at-risk

Exploring the operational risk measurement, just as any kind of risk measurement, must necessarily take us to the scrutiny of a so called Value-at-Risk (VaR). Interested readers might find a good summary of history and evolution of VaR with links to other literature in RiskMetrics (2001), complex theoretical treatment with links to other literature can be found in Jorion (2007). Chernobai et al. (2007, p.211) say on the account of VaR that "it is a powerful statistical tool that has gained popularity in the financial community and has become a benchmark for measuring and forecasting market, credit, operational and other risks." It was made omnipresent especially by the Basel II Capital Accord that established VaR as a benchmark for risk-based internal models for measuring market and credit risk as banks were effectively prescribed to build their quantitative risk management on VaR.

For us denote F_X a cummulative distribution function of a random variable X and recall that it is defined as $F_X(z) = P(X \leq z)$. In the context of the theoretical framework established in previous section we can formally define value-at-risk (for some predetermined time period and distribution of aggreaget losses over this period) in the following way[22]:

Definition 5. Value-at-Risk. *Let X be the random variable representing aggregate losses with a cumulative distribution function F_X. Given $\alpha \in (0; 1)$ the Value-at-Risk at confidence level α, VaR_α, is defined as follows:*

$$VaR_\alpha(X) = inf\ \{z\epsilon\mathbb{R}|F_X(z) \geq \alpha\}.$$

In fact VaR_α is nothing else than the α-quantile of the underlying distribution of aggregate losses (it can be alternatively formulated as $VaR_\alpha(X) = F_X^{-1}(\alpha)$ – see Rockafellar and Uryasev (2000) for details). In our framework VaR at confidence level α is an amount of capital bank needs to keep in order to make the probability of going bankrupt equal to (1- α). Axiom T is clearly satisfied as for $b\epsilon\mathbb{R}$ we have:

$$VaR_\alpha(X - b) = VaR_\alpha(X) - b$$

and hence also $VaR_\alpha\big(X - VaR_\alpha(X)\big) = 0$ (that means that bank can fully offset the risk – satisfy the regulator – by holding capital reserve equal to the value of risk exposure, as our framework requires).

VaR is the dollar (or other currency) value, a single number. Some of its attractive features that make it quite attractive for regulators and managers alike are the ease of calculation, comparability, comprehensiveness and ease of interpretation. In practical applications three parameters need to be set in order to determine VaR – confidence level α, time interval, and base currency (notice that Definition 5 omits the considerations of time and unit of measurement as they are inherent in random variable X as outlined in

[22] Alternative formulations are possible, but the differences between them are of no practical relevance for our research.

previous text). For a given confidence level and for a given timeframe VaR determines the worst possible loss (denominated in the chosen currency) the bank can suffer. Alternative interpretation is that VaR is the smallest possible loss given the worst case scenario (i.e. minimum loss that the bank will suffer in $(1-\alpha)$ percent of worst cases). Basel II Capital Accord requires the operational charge to be computed for periods of 1 year or longer and also at extremely high confidence level of 99.9%.

3.2.1. Value-at-Risk viewed from the coherence paradigm

The concept of coherence brings a strong case against Value-at-Risk. VaR is rejected in Artzner et al. (1999) on the ground of not satisfying subadditivity which is perhaps the most important of the four axioms of coherent risk measures (even though the other three are satisfied). As shown e.g. in Embrechts et al. (2006) or Krause (2003), VaR fails subadditivity especially when losses the losses are extremely heavy-tailed as is typically the case in operational risk modeling, especially with the growing endorsement of EVT methods (a more theoretical treatment of asymptotic properties of VaR under EVT distributions is elaborated in Embrechts et al. (2009)).

Let us consider several stylized examples of violating subadditivity principle and illustrate why we should not rely on such a risk measure (these examples are not directly related to operational risk, but in section 3.4. we explain their relevance for our context).

i) Strikingly simple example is suggested in Albanese (1997). Consider holding two portfolios consisting of two different independent bonds with identical risk-return profiles. Assume that the return on both bonds is 0 with certainty (for simplicity) and that they have identical probability of default of 4% (with no recovery) over the holding period. 95% value-at-risk for each separate portfolio is 0, i.e. VaR assigns no risk, even though each portfolio has a 4% probability of generating loss equal to its nominal value. Now let us consider summing up both portfolio and compute the VaR of the product. since the probability of at least one of the bonds defaulting is now (1-0.96*0.96)=7.84%, joint portfolio would be certainly be assigned a positive VaR (i.e. it would be risky) since the probability of loss now exceeds the chosen confidence level. Diversification of the portfolio created risk while in case of massing the bonds of one company the risk remained completely undetected. Indeed this is an easy example of VaR failing to recognize the concentration of risks and Artzner et al. stress that such a measure should not be relied upon.

ii) Artzner et al. bring an example illustrating how VaR fails to encourage a reasonable allocation of risks among agents, a problem I earlier described as “stuffing the risk into tails" practice often exercised to make the portfolios seem less risky. Consider three possible states of the world, $\omega_1, \omega_2, \omega_3$ with respective probabilities of 0.94, 0.03 and 0.03. Consider agent A and B holding identical portfolios with future net worths $X(\omega_1) > 0$; $X(\omega_2) = X(\omega_3) = -1000$. An extra capital charge of 500 would not be sufficient for neither agent to cover the 5% expected loss computed by VaR. However the two agents might enter an arrangement where their future net worths in ω_2, ω_3 are modified to $X_A(\omega_2) = X_B(\omega_3) = -500$ and $X_B(\omega_2) = X_A(\omega_3) = -1500$ (so now both agents face the loss of 1500 with probability of 3% and loss of 500 with probability of 3%). Because the 5% value at risk of such risk profile is 500, extra capital charge of 500 is now sufficient to make the position acceptable. That is not reasonable since under typical assumption of risk aversion the original position is strictly Pareto-dominating the modified one. This is a simple illustration of a risk measure creating perverse incentives for agents to alter their economic behavior and taking suboptimal position in order to improve the reported risk statistics.

Using the optics of coherence Artzner et al. (1999) summarize two main reasons for rejection of VaR as a risk measure (p.218): “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.”

Artzner et al. (1999) has provoked quite an intensive reaction in the literature. Acerbi (2001a) explains that it was the much-needed and first attempt altogether to introduce a proper theoretical framework for the discussion of risk measures, or as Acerbi puts it all the previous progress was “discussing risk measures without even defining what “risk measure” mean.” Furthermore coherence is defined as a set of axioms and represents what most practitioners and researchers perceive as natural, logical and desirable features *every* risk measure should have. It is far from singling out any “best measure”. But failure to comply with the definition of coherence is necessarily a bad point. In this view coherence is not only a set of “some nice properties”, it is a completely new paradigm allowing to sort out good and bad risk measures or even to say what is a risk measure and what is not.

In the light of coherence the omnipresence of VaR might seem alarming. Indeed coherence as a desirable property has been overlooked in practice for some time despite the numerous responses of the literature. Acerbi (2001b) argues that perhaps because there was no known coherent risk measure that would share the desirable properties of VaR that make it so attractive (simplicity, universality…) many practitioners believed ”that coherence might be some sort of optional property that a risk measure can or cannot display” (p.2) and VaR was being used as much as before with little discomfort caused by its non-coherence. Acerbi has somewhat jokingly depicted the impropriety of such thinking in the following way: “The axioms of coherence simply embody in a synthetic and essential way those features that single out a risk measure in the class of statistics of a portfolio dynamics, just like the axiom “it must be higher when air is hotter” identifies a measure of temperature out of the class of thermodynamical properties of the atmosphere. If you want to use a barometer for measuring temperature despite the fact that pressure does not satisfy the above axiom, don’t be surprised if you happen to be dressed like an Eskimo in a hot cloudy day or to be wearing a swim costume in an icy sunshine” (p.2).

3.2.2. Value-at-Risk and operational risk

We have already explained that the risks X are in context of loss distribution approach to operational risk modeling represented by the aggregate loss distribution which is arrived at by merging fitted frequency and loss distribution of the loss data. These distributional assumptions determine the shape and characteristics of aggregate loss distribution and hence are obviously critical when it comes to application of a risk measure as quantification of risk capital charge. Function of the risk measure then is to take as much information (that we consider important for given purpose) as possible and to process it in a meaningful manner to a single dollar value representing the risk exposure. Ideal risk measure should clearly attain higher values for higher levels of risk, i.e. for riskier loss distributions (in context of operational risk “level of risk” can be perceived e.g. in terms of first-order stochastic dominance in tails of the loss distributions[23]). That is a basic rationale behind coherent risk measures. Unfortunately Value-at-Risk does not always behave this way as we illustrated in previous section on some counterexamples (and lack of sub-additivity is implied precisely by this fact).

VaR can be easily used for determining the capital charges (since it is measured in currency units it is used to calculate the capital charge directly as dollar number as required by Basell II) and provides a uniform metric to compare the risk levels of different loss distributions, but, based on the above-stated properties, it has the following critical drawbacks for operational risk modeling:

1) VaR is only a lower bound for extreme losses. It provides no information on the losses that lie beyond the estimated threshold. Intuitively it is an answer to a misleading question: "What is our minimum loss in 100(1-α)% worst cases?" Since the interest of operational risk modeling focuses on the extreme quantiles of loss distribution this equals to discarding useful information on what we consider as indication of risk exposure (for market risk that is pointed out already in Artzner et al. (1999) or Acerbi and Tasche (2001, 2002), we believe that in operational risk this setback is of even greater importance).

2) VaR may fail the subadditivity property which makes it possibly misleading for the quantification of total risk exposure under AMA of Basel II. Recall that Basel II rules encourage banks to quantify the risk capital charge $\rho(X_i)$ separately for each business unit *i*, and obtain estimate of total risk as:

$$\rho_{BANK} = \sum_{i=1}^{n} X_i.$$

In such case only the sub-additive risk measures provide meaningful results as subadditivity implies that

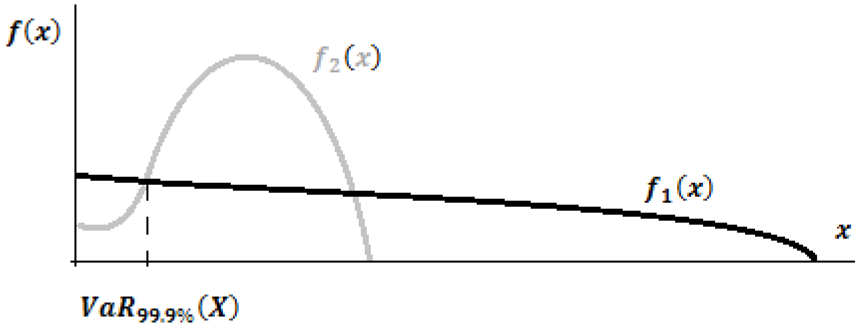
$$\rho_{BANK} = \rho\left(\sum_{i=1}^{n} X_i\right) \leq \sum_{i=1}^{n} \rho(X_i),$$

i.e. that the sum of the separate risks is the upper bound of the total operational risk exposure. In contrast, sum of individual VaRs says nothing at all about the total risk and should not be considered a plausible estimator of risk capital charge under

[23] Stochastic dominance over whole region of losses is of lower importance since we are typically primarily interested in extreme quantiles in context of operational risk modeling.

this approach (for empirical example of VaR failing the subadditivity property see Giacometti et al. (2008)).

Figure 3.1: Examples of two different tail risks.



Source: Author.

3) VaR is suboptimal risk measure for comparisons of tail risks associated with different distributions of losses (which is an implication of point 1), because it does not distinguish between various tail structures of loss distributions behind the α-quantile. As an example consider two stylized PDFs from the Figure 3.1. Function f_1 clearly represents a more significant tail risk (and hence less favorable risk profile) but that goes unnoticed by VaR statistic. It has several serious practical consequences:
    a. It makes the model comparison and selection under AMA using VaR as a measure of risk exposure somewhat dubious – loss distributions with lower VaR statistics can in fact be associated with larger risks which might be "hidden" behind the VaR threshold, in the tails of the distribution (this point is clearly demonstrated in empirical part of our research).
    b. Since operational risk management is a continuous activity one needs to constantly update the models (e.g. with arrival of new observations of operational losses). Again, intermodel comparison using VaR can conceal the changes in risk exposure if these are induced by shifts in the tail structure of the loss distribution.
    c. Finally from institutional viewpoint, as advised by Yamai and Yoshiba (2002), VaR-based reporting can fail to provide incentives for purposeful risk management as precautions directed to mitigation or reduction of severity of extreme events might not be reflected in the VaR statistic (e.g. when these are considered in scenario analysis VaR calculations).

Note that some authors are suggesting „alternative“ methods to VaR defined as quantiles of a right tail aggregate loss distribution, e.g. Median Tail Loss measure (tail is meant here as the distribution of losses behind some VaR_α threshold). But in fact that is nothing else than the VaR of a higher confidence level. All objections made against VaR can be made with same degree of insistence against these measures as well. Even though they provide more conservative capital estimates than the VaR of the same confidence level their defect is principally the same – they are discarding some information from the rightmost part of the distribution neglecting possible extreme events.

3.3. Representation of coherent risk measures: Expected Shortfall

Seminal paper of Artzner et al. (1999) has caused quite a vivid discussion on the usefulness of VaR and has triggered many attempts to upgrade VaR to comply with the subadditivity property and hence qualify for coherence. One of these attempts has been briefly suggested already in the mentioned paper resulting in the statistic known as Expected Shortfall. To avoid confusion, this method is also called Conditional VaR, Mean Excess Loss and also Tail VaR, as many authors have further elaborated upon this idea – some of the early studies presenting the desirable properties of ES include Pflug (2000), Rockafellar and Uryasev (2000), Acerbi et al. (2001), Acerbi and Tasche (2001, 2002), Hurliman (2003) or Inui and Kajima (2005).

As a motivation to derivation of ES, let us consider the following question: *“What is our expected loss in 100(1-α)% cases?”* When one is interested in the estimation of extreme losses such question definitely makes more sense than the one asked by VaR and it immediately evokes the following statistic:

Definition 6. Upper Conditional VaR. *Let X be random variable representing aggregate losses. At the chosen confidence level* $\alpha \in (0,1)$ *the Upper Conditional VaR,* $CVaR_X^+$*, is the following statistic:*

$$CVaR_\alpha^+(X) = E[X|X > VaR_\alpha(X)].$$

Upper CVaR gives the mean of losses X strictly exceeding the VaR threshold for some confidence level α. To avoid confusion we call this statistic “Upper Conditional VaR” following Hurliman (2003) and Sarykalin et al. (2008). This statistic is useful for understanding ES, nevertheless it is not a coherent risk measure for general distributions of X. Specifically it is only coherent when X is continuous, which is absolutely not a plausible assumption in case of aggregate operational loss distributions. Yet by the proposition in Rockafellar and Uryasev (2002) ES can be obtained as a weighted average of $CVaR_\alpha^+(X)$ and $VaR_\alpha(X)$ and it is a coherent measure in sense of Artzner et al. (1999):

Definition 7. Expected Shortfall. *Let X be a random variable representing aggregate losses with cumulative distribution function* F_X. *At the chosen confidence level* $\alpha \in (0,1)$ *Expected Shortfall is defined as:*

$$ES_\alpha(X) = \frac{F_X\big(VaR_\alpha(X)\big) - \alpha}{1-\alpha} VaR_\alpha(X) + \frac{1 - F_X\big(VaR_\alpha(X)\big)}{1-\alpha} CVaR_\alpha^+(X).$$

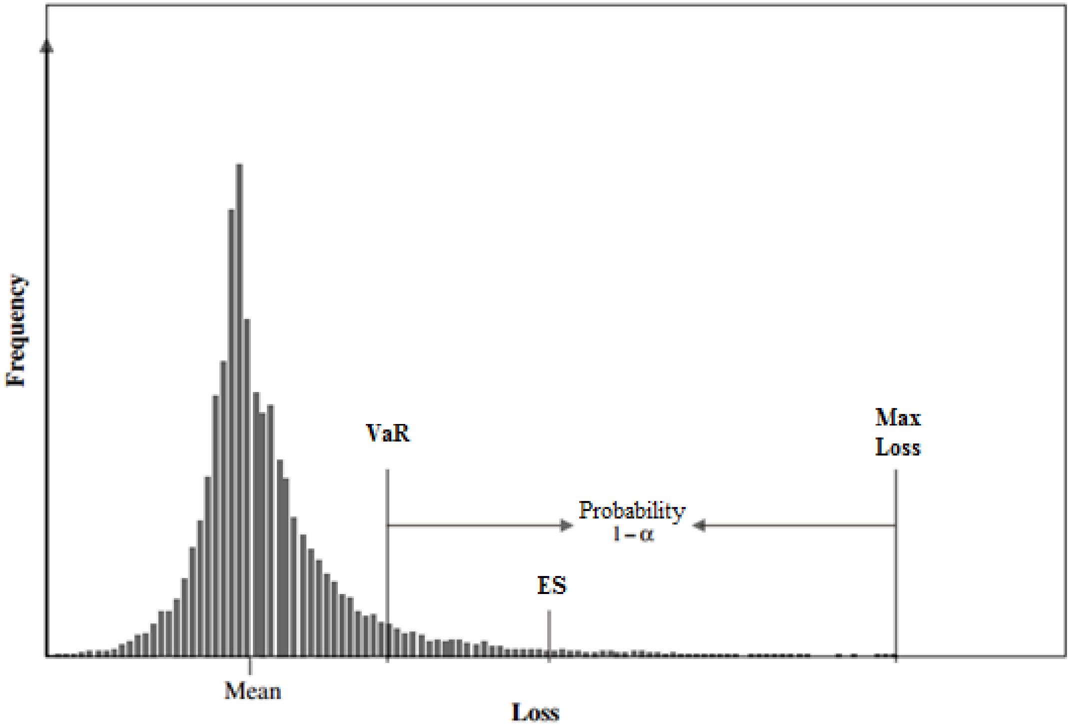
The difference between definition 6 and 7 embodies an adjustment needed to ensure the coherence of the statistic under non-continuous distributions[24] (for continuous distributions the weights are obviously equal to 0 and 1 for VaR and Upper CVaR respectively, hence the formula reduces to one in Definition 6). Hurliman (2003) presents a list of alternative formulas defining ES, we prefer this one and use it for calculations in empirical part for its simple applicability. Expected Shortfall can be thought of as an average measure of heaviness of the tail of the distribution or an average worst-case loss adjusted for the discontinuity of the distribution function (stylized illustration of VaR and ES is given in Figure 3.2). Expected Shortfall has been proved to be sub-additive in Pflug (2000) or Acerbi and Tasche (2001, Appendix A). Some detailed comparative analysis of VaR and Expected Shortfall (though mainly for market risk or simulated data) can be found e.g. in Yamai (2002), Korkhmal et al. (2001), Heyde et al. (2007), Sarykalin et al. (2008).

ES, contrary to VaR, evaluates risk exposure using whole tail distribution including the most extreme losses. It is hence a very straightforward and natural measure of tail risk, furthermore it possesses amenities of coherent risk measures, such as subadditivity, and also shares some of VaR's attractive features – it is intuitive and easy to implement. It is an answer to all three main problems stated on account of VaR in the previous section.

[24] Hence as Sarykalin (2008, p.273) hints „contrary to popular belief, in the general case CVaR_α (*ES in our notation – author*) is not equal to an average of outcomes greater than VaR_α." This paper can be viewed for more intuitive insight to ES derivation.

Naturally there are also some caveats related to this statistic – mainly since it uses more information from the loss distribution it tends to have worse convergence properties and larger variance than VaR of the same confidence level (as can be seen in Yamai and Yoshiba (2002) or Broda and Paolella (2009)).

Figure 3.2: VaR and ES for stylized loss distribution



Source: Author based on Sarykalin et al. (2008)

Aalthough the use of simple VaR is still predominant in the field of operational risk modeling ES has recently been considered in several studies and proposed as a viable alternative for calculation of capital requirements. Biagini and Ulmer (2008), Lee and Fang (2010) or Giacometti et al. (2008) are the examples of recent operational risk research drawing upon the concept of Expected Shortfall. Aue and Kalkbrener (2006) convey that it is also a consistent tool for capital allocation purposes. A variant of Expected Shortfall is also considered in Moscadelli (2004).

3.3.1. Comparing Expected Shortfall and Value-at-Risk

Guegan and Tarrant (2011) admonish that in order to realy understand the risks related to given loss distribution one needs to use several[25] different risk measures in a complementary fashion. Consistently with this idea we are convinced that useful information can be obtained on the loss distribution by comparison of VaR and ES statistics. Namely it gives us the idea on the heaviness of the tail, an answer to a metaphorical question: "How big is the black swan lurking behind the VaR threshold?"

We suggest two different ways of comparison that can be thought of as crude measures of heaviness of the tail. First is a simple ratio of ES/VaR. This is by definition of the two statistics always greater than 1. Higher ratio value is associated with heavier tail (more severe worst-case losses behind the VaR level). Hence the riskier of the two distributions with the same VaR can be distinguished by higher ES/VaR statistic. We call our second comparison method "Equivalent α":

Definition 8. Equivalent α. *Let X be a random variable representing aggregate losses and $\bar{\alpha} \epsilon (0,1)$ a predetermined confidence level. Equivalent α, denoted as α_e, $0 < \alpha_e \leq \bar{\alpha}$, is a number satisfying the following property:*

$$ES_{\alpha_e}(X) = VaR_{\bar{\alpha}}(X).$$

Existence of Equivalent α is ensured by one of the characteristic and attractive features of ES. Let us for a while consider the discussed risk measures as functions of α. Obviously, when X is not continuous, i.e. there are "vertical gaps" in $F_X(x)$, then $VaR_\alpha(X)$ and $CVaR_\alpha^+(X)$ are not continuous in α (i.e. there are intervals of confidence levels for which these statistics are constant). On the other hand, ES is continuous with respect to α for general distributions as shown in Sarykalin et al. (2008). Hence for any $0 < x \leq ES_{\bar{\alpha}}$ we can find some α_e such that $ES_{\alpha_e} = x$, which implies existence of Equivalent α.

Equivalent α gives us a conception of VaR's underestimation of risks. What VaR reports as a minimum loss in $100(1-\bar{\alpha})$% cases is in fact an average loss in 100(1-α_e)% cases. In this case lower value of the statistic implies larger risk associated with the tail of

[25] Authors state that at least five different risk measures are needed for exact description of the tail of the distribution.

underlying loss distribution. A problem with this statistic is that it is incomparable across different confidence levels of VaR and hence harder to interpret. On the other hand it allows for intermodel comparisons at identical confidence level and we further believe that it can be used as a basis for determining an "appropriate confidence level" for calculation of regulatory capital charge should ES be used for this purpose (as under Basel II such approach would benefit from the advantages of ES while not ruining the consistency of Advanced Measurement Approaches by delivering disproportionally larger capital estimates than VaR-based calculations).

In empirical part of the research we report both these comparative statistics along with ES and VaR estimates and demonstrate how they can be used in quantitative operational risk management.

3.3.2. Expected Shortfall and economic capital

In our view, ES for some sensibly chosen confidence level is an eligible measure of economic capital charge. It is a measure describing the extreme quantiles of loss distributions, likewise VaR, and furthermore it accounts even for the most severe losses. Bearing in mind the objective of economic capital - that is to protect the institution from insolvency due to large unexpected losses - ES is answering an apt question: "What is an average magnitude of the worst case loss?" Even though in ideal case the economic capital should cover even the maximum possible loss such prudence is typically not feasible for two reasons:

1) *We never know the maximum loss with certainty. Under the Loss Distribution Approach we can at best get a point estimate but it will be different simulation form simulation.*
2) *Corresponding level of capital could be prohibitively large.*

No matter how conservative the measure is some residual risk will always remain. Except for some abnormal cases, ES is always less than the maximum loss amount hence even when ES is used as a measure of economic capital we still need to keep in mind that catastrophic events might occur for which the capital buffer might be insufficient. But having the capital buffer set at ES level bank can expect to be able to cover the average worst case loss (as opposed to minimum worst case loss in case of VaR).

3.4. Criticism and discussion of coherence

Not only VaR was subject to criticism. Gourieroux and Jasiek (2010) in their defense of VaR argue that “if the VaR was replaced by a coherent risk measure by the regulators, banks would be motivated to merge in order to diminish the amount of required capital (subadditivity axiom). Clearly such an incentive to merge may create non-competitive effects and increase the risk”. While an interesting idea it seems somewhat narrow-sighted. Goal of efficient risk management is to find ways to properly control the risks and that might only be done by using tools that are able to give as realistic a picture of our risk profile as possible. If proper risk management and implied prudence measures provide additional incentives for mergers then be it. Risk measures are out to help quantify risks – what is to be done with such information is a matter of managerial decision (but indeed the resulting action can be subject to operational risk considerations too). Besides the subadditivity principle only encourages mergers to a degree of dissimilarity of the agents’ portfolios and investment strategies while at the same time these dissimilarities might increase the merger costs.

RiskMetrics (2001) goes a different way. Authors do not doubt the conceptual logic of subadditivity principle; they instead impugn its practical importance: “While VaR does not always satisfy the subadditivity property, it is difficult to come up in practice with cases where it is not satisfied.” Yet we have seen many examples where failing to comply with this property imposed severe practical implications, in fact we could go on inventing countless new examples where non-additive risk measures produce result that obviously wrong – and indeed dozens of such have already been provided in the literature - as a breach of the axiom creates somewhat paradoxical state of affairs. It might really be the subadditivity property what is the characteristic attribute of the very concept of risk – diversification principle is perhaps the first and most important principle of risk management and as such it should be reflected by any attempt for quantification of risk, unless we want to have a wide gap between theoretical literature and the real-world application. Besides some empirical evidence with operational risk data on super-additivity of VaR exists[26], e.g. in Giacometti et al. (2008).

[26] We unfortunately cannot analyze this problem on our dataset due to the fact that it is not sufficiently large to allow for decomposition by business units or event types.

There is one more issue which one often encounters in actuarial literature and which illustrates that the property of subadditivity is not always desirable – specifically in the case where the two merged risks are highly dependent. In insurance practice (see e.g. Dowd (2006) or Dhaene et al. (2003)) it is customary to charge more than double premium for insuring two houses in the same neighborhood compared to premium for an insurance of a single house because the risk is supposedly more than twice as high – here the subadditivity principle clearly contrasts with the best practice. It is a matter of importance to assess the relevance of such affirmations for risk management in banking sector and specifically for operational risk management. Nevertheless with the diversity and complexity of tasks raised by quantitative risk management we can hardly expect to find a risk measure that would be ideal for each and every practical purpose at hand – if subadditivity is a problematic assumption for certain specific cases in insurance risk measurement it is indeed an evidence of a complexity of risk management science, but it still does not disqualify the relevance of subadditive risk measures when the principle of subadditivity is logical and natural in all of the other - or most of the other - cases.

Dhaene (2006) further notes that the terminology "coherent" can be somewhat misleading in this sense as it might suggest that any risk measure that is not coherent is always inadequate. He calls attention to the obvious fact that the properties that a risk measure should satisfy depend on the risk preferences in the economic environment under consideration. It is true that from normative point of view the "best set of axioms" is non-existent because any normative axiomatic setting is based on the "belief" in the axioms – and that is naturally subjective and situation-specific and might for instance represent certain school of thought. But in our view the choice of the term "coherent" was not only strategic, but also made perfect sense – the article did not propose just a new alternative axiomatic setting for risk measures, it proposed *the first* axiomatic setting and its name was not an attempt to discredit the alternative sets of axioms as there were none at the time.

Some of the criticisms surely are reasonable but again their validity must be judged (as well as validity of proposed axioms of coherence) from the standpoint of the specific purpose at hand. It is only proper to conclude that it is of crucial importance that risk managers do not dismiss their common sense and evaluate individual risk measures with respect to the specific context they are facing, instead of accepting the universal truths. It cannot be generally said that one risk measure is better than another but we can search for the risk measures with theoretical properties that we found desirable in the context of operational risk, acknowledging that our findings might not hold in full generality for all branches of risk modeling.

We have explained that the core of quantitative operational risk management lies in aggregation of data and in the estimation of impact of extreme events and in this light the advantages of a coherent method seem immense – as explained earlier coherence allows for safer aggregation of data (e.g. across business lines) and at the same time removes another drawback of VaR by taking whole tails of distributions into consideration. In the context of operational risk modeling the specific properties of coherent risk measures are highly desirable and address the specific problems stemming from usage of VaR.

Chapter 4. Empirical results

4.1. Methodological Remarks

This part of the thesis exploits the theory described in previous sections to model the operational risk exposure of a real commercial bank. In section 4.2 we give a short review of the dataset. Then we continue with fitting various statistical distributions to the data and evaluate their suitability using the goodness of fit tests outlined Chapter 2. Namely we show the QQ plot and outcomes of Kolmogorov-Smirnov and Anderson-Darling tests (it is proper to remind the reader that their null hypothesis is that the data follow the fitted distribution). Aim of this exercise is to find a distribution that is able to model the severity aspect of the data reasonably well so that later we can use it to model the aggregate operational losses and calculate risk exposure using the risk measures. While QQ plot is only a visual inspection tool the two tests are formal comparisons of the fitted and empirical distribution functions. Since the tests might have different critical values depending on the underlying distributions we are only indicating the p-values (so that for any significance level beyond the stated p-value the null gets rejected and we cannot say that the data follow the fitted distribution function). Also note that in order to honor the confidentiality of the data the quantile values are rescaled relative to the largest loss in the dataset.

Critical part of the results is the risk capital charge calculated for each of the distributions. We are calculating Value-at-Risk and Expected Shortfall at 95%, 99% and 99.9% confidence levels. All capital charges are reported as percentages of the gross income ratio from Basel II. VaR is reported for two reasons – it serves as a benchmark measure but also the 99.9% VaR coincides with the regulatory capital charge required by Basel II, as discussed above. ES is a representation of coherent risk measures that suggested in this thesis as a viable alternative (or at least a complement) to VaR and as a good tool for evaluation of the economic risk capital charge. Alongside, two statistics are reported to give a quick comparison of VaR and ES. First for each confidence level the ratio of ES/VaR (which is by definition of the two always greater than 100%) is shown and finally we calculate an "equivalent α" which is a confidence level such that for some α' we have $VaR_{\alpha'}(X) = ES_{\alpha}(X)$. These two statistics give an indication of the heaviness of the tail and allow a risk manager to distinguish the riskier of the two seemingly identical distributions of losses (such information can be extremely valuable in the case when the body of the distributions are similar but one has a heavier tail, hence larger risk). It is our indication of how much trouble lies behind the VaR threshold.

Regarding the methodology suggested in Basel II, one remark is in order. The regulatory framework recommends that in ideal case the risk exposure should be calculated for each business line (to account for their varying vulnerability to operational risk) and the total risk exposure of the bank then obtained as a sum of individual risks across business lines. Unfortunately such approach is absolutely impossible given the limited size of the dataset and indeed when using internal data this is a pervasive problem that many researchers are facing in operational risk modeling[27]. Some of the statistical methods we are employing, such as EVT, would in fact ideally require a much larger dataset than the one we have at our disposal, dividing it between eight business lines would most likely yield unreliable and possibly misleading results.

4.2. Initial Data Analysis

I am using an internal dataset of a Central European commercial bank, further denoted as "BANK", which contains its operational losses over the span of more than three years. Total number of observations in the dataset is 633. Even though the data is truncated from the left, BANK is using an unusually low (practically negligible) threshold for recording the operational losses hence the corrections for truncation bias seem unnecessary. For all the calculations we are ignoring impact of exchange rates fluctuations and only consider the loss amount in the base currency.

To reveal some of the important data characteristics we have selected some basic statistics and summarized them in Table 4.1. The distribution is also heavily right-skewed

[27] That does not at all impeach our argumentation from Chapter 3 where we asserted that under such approach sub-additivity of coherence risk measures is an essentially needed property for ensuring the plausibility of results, it merely means that this argument is of no practical relevance for our illustrational empirical exercise.

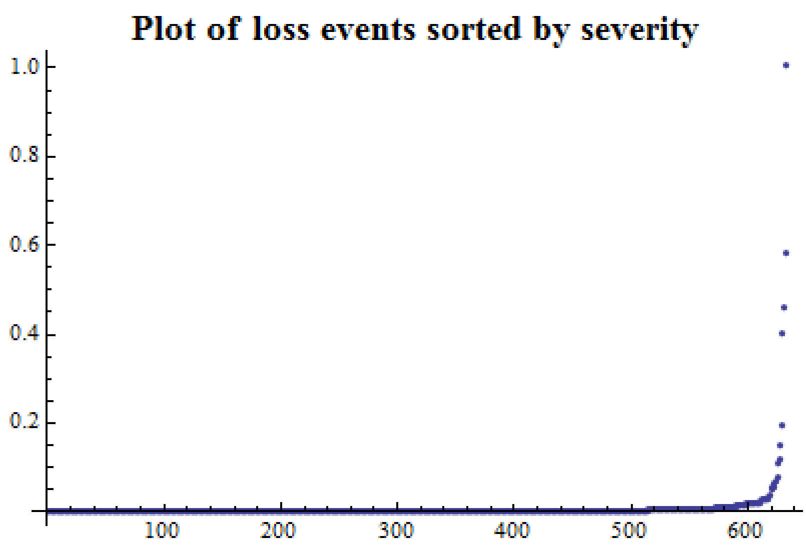
with extremely high kurtosis. The difference between the mean and median (both of these values are scaled by the largest loss in the dataset) is also striking. Mean is nearly 14 times higher than median and is in fact close to 93rd percentile of the empirical distribution function, which is signaling heavy tails. Figure 4.1 and Figure 4.2 give some further insight. First figure displays the losses sorted by severity (one can easily observe the huge difference between the highest losses and the rest of the losses, e.g. only 4 of total 633 observations exceed 20% of maximum loss amount). Figure 4.2 shows individual losses in the order of appearance (dots) and their cumulative sum (line). All this information immediately leads to impression that conventional statistical distributions will have problems to fit whole data sample and that we will have to employ some heavy-tailed distributions to model the data, which is further confirmed.

Table 4.1: Some statistics of the data sample (scaled by largest loss in the dataset)

Mean	***Median***	***Mean/Median***	***Skewness***	***Kurtosis***
0.76%	0.06%	1377%	13.88	220.85

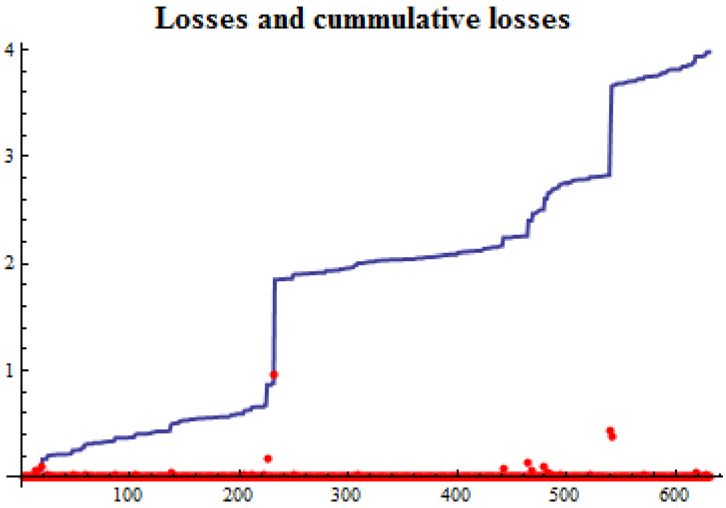
Source: Author.

Figure 4.1: Plot of loss events sorted by severity (scaled by maximum loss amount)



Source:Author.

Figure 4.2: Plot of losses and cumulative losses in order of appearance (scaled by maximum loss amount)



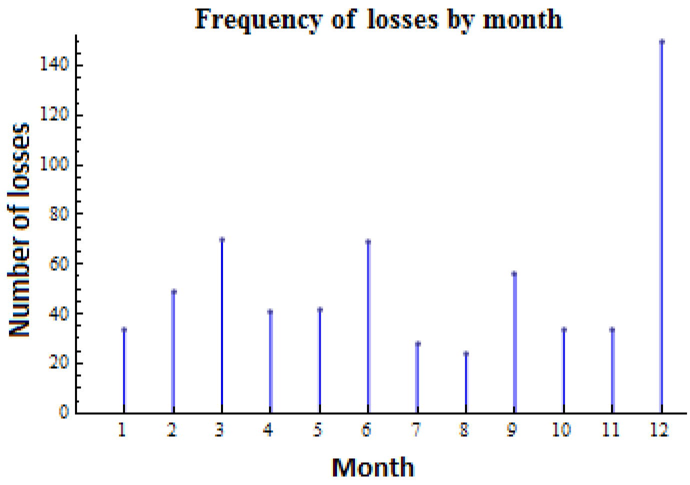
Source: Author.

4.3. Fitting frequency distribution

Many researchers agree that frequency aspect of the data is not as critical as severity and modeling of frequency by Poisson process has become a standard in operational risk research. Even though a better result might have possibly been reached using some more sophisticated method, Poisson distribution is still able to model the data sufficiently well and is perfectly acceptable for our purposes, hence it is used for all our applications as a model of frequency.

One slight complication is revealed in Figure 4.3 where one can easily observe that number of events is fairly different across the months. Typically for operational risk data there is a strong seasonal effect in December as the losses are recorded at the end of accounting period, but also for other months the number of events is varying considerably. Hence modeling the frequency as a single Poisson process would provide a poor fit to the data (as a visual tool to evaluate the fit of single Poisson process is to see whether the number of events in Figure 4.3 roughly follows a horizontal line, but it clearly does not). To settle this problem and to improve the model slightly we model the frequency for each month separately obtaining the maximum likelihood estimator of intensity rate $\hat{\lambda}$ simply as yearly average of number of loss events in a given month.

Figure 4.3: Loss frequency by month of appearance



Source: Author.

4.4. Simulating loss severity with empirical distribution

This section presents the results of Monte Carlo simulation where empirical distribution is used as the distribution for loss severity. This method is sometimes referred to as empirical sampling method (ESM). ESM does not represent a proper model of aggregate loss distribution and would most likely not be accepted by the regulator as a valid AMA approach to operational risk modeling. But for two reasons we find it useful to perform this exercise – first it is to serve as a benchmark for comparison of other methods and second, ESM is particularly convenient for modeling of body of the distribution when the tail is modeled by some method from Extreme Value Theory.

Estimated capital charge is very low. If the bank was using Standardized Approach from Basel II for assessing the operational risk capital it would need to hold 15% of the calculated Basel gross income ratio, whereas this method only prescribes 2.45% as regulatory capital charge or 2.66% as economic capital charge (Table 4.2). Such result could in fact be expected. A reasonable model of aggregate operational losses should generate capital charges higher than this simple approach. Reason is simple – each random variate selects a random loss from the dataset with equal probability (hence most of the losses will be very low, close to the data recording threshold) and furthermore the losses drawn from empirical distribution are bounded from above by the maximum loss in the sample. No matter how many random variates we draw single loss will never exceed this amount. In case of fitted parametric distributions this will not be the case (only in the specific case of beta distribution, where the losses will also be bounded from above by the maximum loss, see section 4.5.5 for details).

Table 4.2: Simulated risk capital charges for empirical distribution of loss severity

Empirical Distribution				
α	***ES***	***VaR***	***ES/VaR***	***Equivalent* α**
99.9%	2.66%	2.45%	108.52%	99.73%
99%	2.15%	1.90%	112.86%	97.36%
95%	1.73%	1.46%	118.61%	13.52%

Source: Author.

4.5. Fitting the severity distribution

In this section we try to model the severity of losses using several well-known statistical distributions that have been considered in the context of quantitative risk modeling. Even though the fit is generally poor we nevertheless calculate the aggregate loss distribution and corresponding capital charges to illustrate the behavior of VaR and ES. We are using the Monte Carlo simulation engine for aggregation as described in section 2.5. To give a short summary, we first draw 12 random variates from Poisson distributions fitted for each month in a year and sum them up obtain the simulated number of operational events over the span of one year (denote N). Then we draw N random numbers from the fitted severity distribution and sum them to obtain simulated aggregate operational loss for one year. We repeat this process is repeated 100 000 times to obtain a numerical approximation of the distribution of aggregate losses (with 100 000 "observations") and finally we apply various risk measures to estimate the risk exposure of a bank.

Our selection of distributions includes exponential, lognormal, Weibull, gamma, beta, Pareto distribution and g-and-h distribution, as they have been presented in Chapter 2. Note that for each distribution two QQ plots are shown. Plot on the left displays overall fit with all possible quantiles (i.e. one for every observation in the dataset) while right-hand side plot shows only 90% of lower quantiles, thus it represents the fit in the body of the distribution only. All quantiles are scaled by the largest loss in the dataset, again.

Last remaining point we would like to address here is the question how to select an appropriate method. We examine several aspects of the results similarly to Dutta and Perry (2007). First obvious criterion is the goodness of fit of the distribution on the underlying data. Secondly, if a method fits well to the data in statistical sense, we regard whether it also produces an aggregate loss distribution with realistic capital estimates (these first two aspects are given most attention in section 4.5 and 4.6). Another important aspect is flexibility, i.e. how can the method cope with a wide variety of empirical loss data (for some selected methods this question is addressed in Section 4.8). Last and perhaps least important criterion (but possibly more relevant in practice than in the theoretical research) is the ease of implementation, where one can inquiry e.g. how complicated it is to obtain parameter estimates and generate random variates from the given distribution.

4.5.1. Exponential Distribution

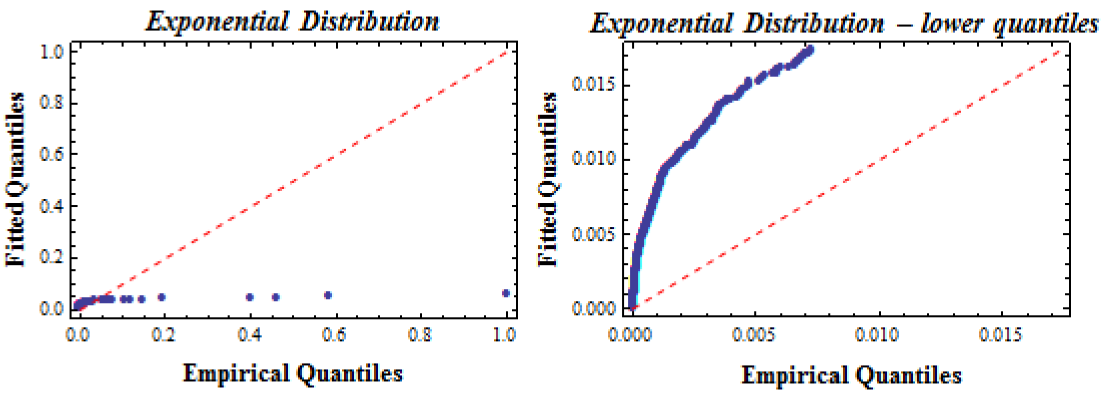
Table 4.3: Estimated parameters and GOFTs for exponential distribution

Exponential Distribution	
$\hat{\lambda}$	8.7751*10^-7
KS (p-value)	<0.01
AD (p-value)	<0.01

Source: Author.

Exponential distribution has an exponential decay in the tail and, using the terminology of Chernobai et al. (2007) it is a typical “thin-tailed” distribution. Such distributions should be used with caution in operational risk modeling as they might severely underestimate the probability of extreme events. Left panel of Figure 4.4 illustrates this quite clearly - we can see a clearly downward sloping plot curve signalizing a gross underestimation of the highest quantiles (the last plotted empirical quantile is more than 10 times higher than its fitted counterpart). Per contra, the lower quantiles are overestimated as the right panel reveals. Hence exponential distribution has a poor fit also for the body of the distribution, which makes it practically useless for modeling data of similar kind. Also the null hypothesis of Kolmogorov-Smirnov and Anderson-Darling tests can be rejected on any reasonable significance level.

Figure 4.4:QQ plots for the fitted exponential distribution



Source: Author.

Table 4.4 shows two interesting pieces of information. First it confirms our conclusion that the risk exposure is underestimated by exponential distribution because on 99.9% significance level the calculated capital charge is near to 1% of the gross income (compared to 15% prescribed by Standardized Approach that is indeed very little). Second, the above statement is true for both Value-at-Risk and Expected Shortfall statistics – it is a good illustration of the above-mentioned assertion that for thin-tailed distributions the two are not drastically different. The difference increases slightly with lower confidence levels, but both measures are still perfectly comparable in this case due to very fast decay of the aggregate loss distribution beyond the VaR thresholds with high α.

Table 4.4: Simulated risk capital charges for exponential distribution

Exponential Distribution				
α	***ES***	***VaR***	***ES/VaR***	***Equivalent α***
99.9%	1.00%	0.98%	102.45%	99.76%
99%	0.93%	0.90%	103.48%	97.36%
95%	0.88%	0.84%	104.25%	87.42%

Source: Author.

4.5.2. Lognormal Distribution

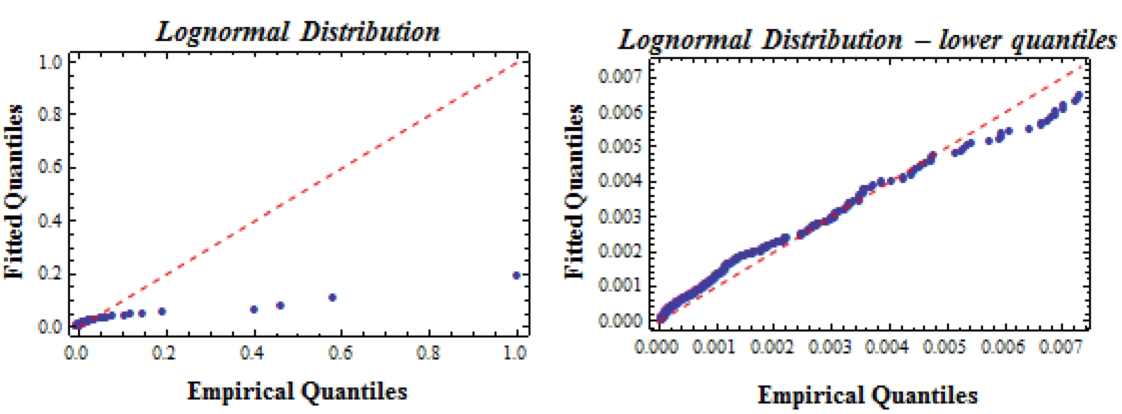
Table 4.5: Estimated parameters and GOFTs for lognormal distribution

Lognormal Distribution	
$\hat{\mu}$	11.5246
$\hat{\sigma}$	1.77525
KS (p-value)	<0.01
AD (p-value)	<0.01

Source: Author.

If we concentrated solely on KS, AD statistics and the left panel of Figure 4.5 we would conclude that the fit of lognormal distribution is just as bad as that of exponential distribution. Highest quantiles are still heavily underestimated, even though somewhat less than in previous case. A little heavier tail is suggested also in the risk exposure calculation where we can now observe the difference between VaR and ES (it is most distinctive at 99.9% significance level where ES is almost 50% higher than VaR). Calculated capital charges are still very low, even lower than for ESM method, as in exponential case. However observing the QQ plot of lower quantiles we can see that the fit in the body of the distribution is in fact quite accurate. Problem is again in the heavy tail of the empirical distribution function, which cannot be accurately modeled by this distribution (even though Chernobai et al. (2007) classify it as heavy-tailed). We can conclude that lognormal distribution is not suitable for modeling operational losses, nevertheless we could still consider it at least for modeling of the body of the loss distribution, e.g. when we need it for the purposes of Monte Carlo simulations under Extreme Value Theory approach (as will be discussed later).

Figure 4.5: QQ plots for the fitted lognormal distribution



Source: Author.

Table 4.6: Simulated risk capital charges for lognormal distribution

Lognormal Distribution				
α	***ES***	***VaR***	***ES/VaR***	***Equivalent* α**
99.9%	1.84%	1.25%	147.45%	99.66%
99%	0.93%	0.68%	136.02%	96.54%
95%	0.63%	0.49%	128.79%	83.78%

Source: Author.

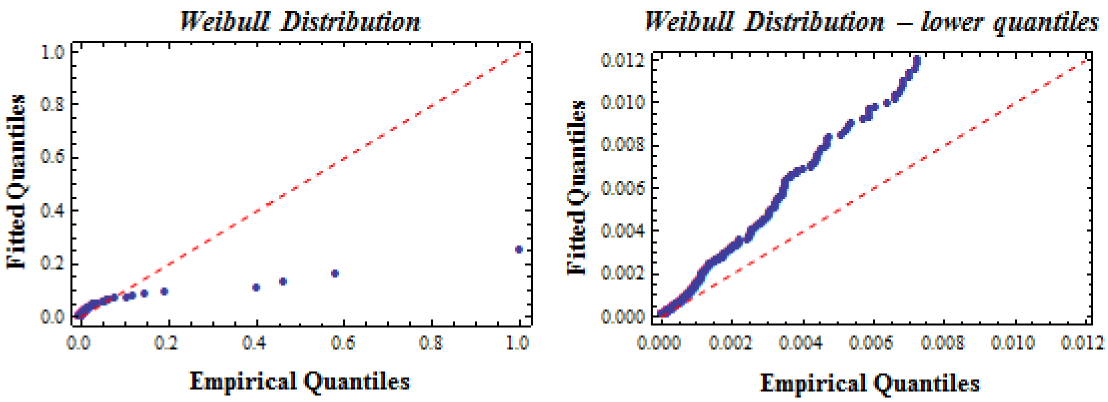
4.5.3. Weibull Distribution

Table 4.7: Estimated parameters and GOFTs for Weibull distribution

Weibull Distribution	
$\hat{\tau}$	0.63706
$\hat{\beta}$	245940
KS (p-value)	<0.01
AD (p-value)	<0.01

Source: Author.

Figure 4.6: QQ plots for the fitted Weibull distribution



Source: Author.

Depending on the value of parameter τ, Weibull distribution can be classified both as thin-tailed ($\tau > 1$, according to Chernobai et al. (2007)) or as heavy-tailed distribution ($\tau < 1$). Hence a sole estimation of τ parameter can give some useful insight to the characteristics of underlying data. In our case we obtained $\tau = 0.64$, which is in line with our expectation. Nevertheless the overall fit is still very poor with lower quantiles overestimated and higher quantiles severely underestimated. Hypothesis that our operational loss data follow the Weibull distribution is denied also by extremely large test statistics of both formal tests of goodness of fit, particularly in AD case which puts more emphasis on the extreme quantiles. Certainly, definition of Weibull distribution allows for much heavier tails than this. Estimate of the shape parameter τ is still too low to fit the right tail, perhaps owing to the MLE estimation method. With other method that would put more emphasis on extreme quantiles we might have had achieved a better fit in the tails, on the other hand the lower quantiles are overestimated already with these parameter values, so there is no reason to believe that Weibull distribution of any specification could reasonably model the data.

Implied capital charges, as summarized in Table 4.8, are unrealistically low and resemble the thin-tailed distributions described above. There is only a moderate difference between ES and VaR and a slow decline of both statistics for decreasing confidence level suggests that only small number (if any) of extreme aggregate losses has been generated.

Table 4.8: Simulated risk capital charges for Weibull distribution

Weibull Distribution				
α	***ES***	***VaR***	***ES/VaR***	***Equivalent*** **α**
99.9%	1.65%	1.32%	125.25%	99.74%
99%	1.18%	1.05%	112.38%	97.14%
95%	0.98%	0.87%	112.64%	86.98%

Source: Author.

4.5.4. Gamma Distribution

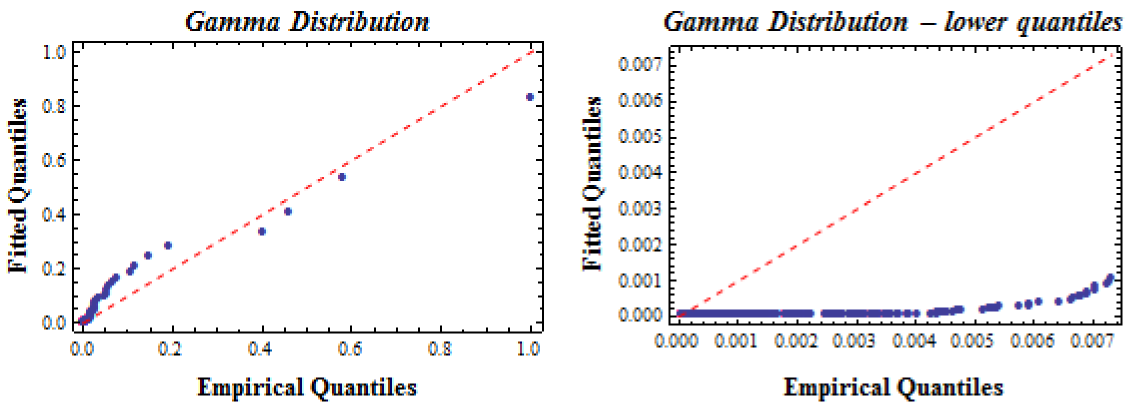
Table 4.9: Estimated parameters and GOFTs for gamma distribution

Gamma Distribution	
$\hat{\alpha}$	0.0205694
$\hat{\beta}$	55401600
KS (p-value)	<0.01
AD (p-value)	<0.01

Source: Author.

All goodness of fit measures reject the hypothesis that gamma distribution provides a good model for our data. KS and AD tests reject the null at any thinkable confidence level. The QQ plots exhibits a solid fit for the most extreme right quantiles, but by contrast, the fit in rest of the distribution is poor – lowest quantiles are systematically underestimated (right panel of Figure 4.7) while the rightmost approximately 10% quantiles (except for the most extreme ones, which fit quite well) are systematically overestimated. Despite these problems the overall provides the most accurate picture so far and also the calculated capital charges are somewhat more plausible (but still too low compared to 15% from the Basel II Standardized Approach).

Figure 4.7: QQ plots for the fitted gamma distribution



Source: Author.

Table 4.10: Simulated risk capital charges for gamma distribution

Gamma Distribution				
α	***ES***	***VaR***	***ES/VaR***	***Equivalent* α**
99.9%	2.81%	2.55%	110.11%	99.75%
99%	2.17%	1.89%	114.83%	97.31%
95%	1.72%	1.43%	119.66%	86.73%

Source: Author.

4.5.5. Beta Distribution

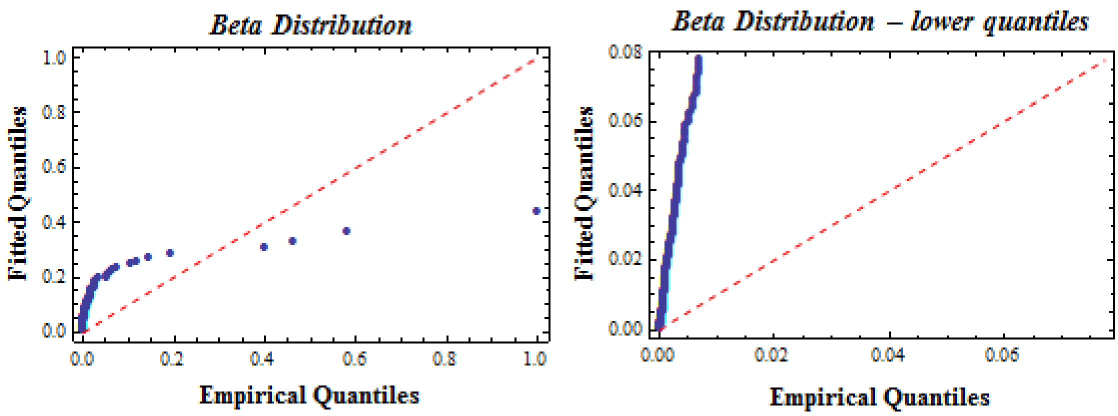
Table 4.11: Estimated parameters and GOFTs for beta distribution

Beta Distribution	
$\hat{\alpha}$	0.186659
$\hat{\beta}$	5.47395
KS (p-value)	<0.01
AD (p-value)	<0.01

Source: Author.

Since beta distribution has a bounded support, to find estimates of the parameters we first divided the original data by the maximum loss amount to obtain rescaled data in the interval [0,1] (see section 2.3. for details). QQ plot exhibits a clear overestimation of all but the most extreme quantiles which are, per contra, underestimated. Calculated capital charges seem reasonable regarding the total amount, but given the ill fit we cannot conclude that the aggregate loss distribution generated with beta distribution represents a good model of operational risk exposure of the bank (estimated capital is relatively high due to overestimation of low quantiles). Negligible difference between ES and VaR also reveals that aggregate loss distribution does not have a particularly heavy tail and quickly converges to zero behind the VaR threshold.

Figure 4.8: QQ plots for the fitted beta distribution



Source: Author.

Table 4.12: Simulated risk capital charges for beta distribution

Beta Distribution				
α	***ES***	***VaR***	***ES/VaR***	***Equivalent* α**
99.9%	5.31%	5.09%	104.20%	99.75%
99%	4.75%	4.52%	105.11%	97.41%
95%	4.35%	4.08%	106.70%	87.38%

Source: Author.

4.5.6. Pareto Distribution

Table 4.13: Estimated parameters and GOFTs for Pareto distribution

Pareto Distribution	
$\hat{\alpha}$	0.79263
$\hat{\beta}$	60918
KS (p-value)	<0.01
AD (p-value)	<0.01

Source: Author.

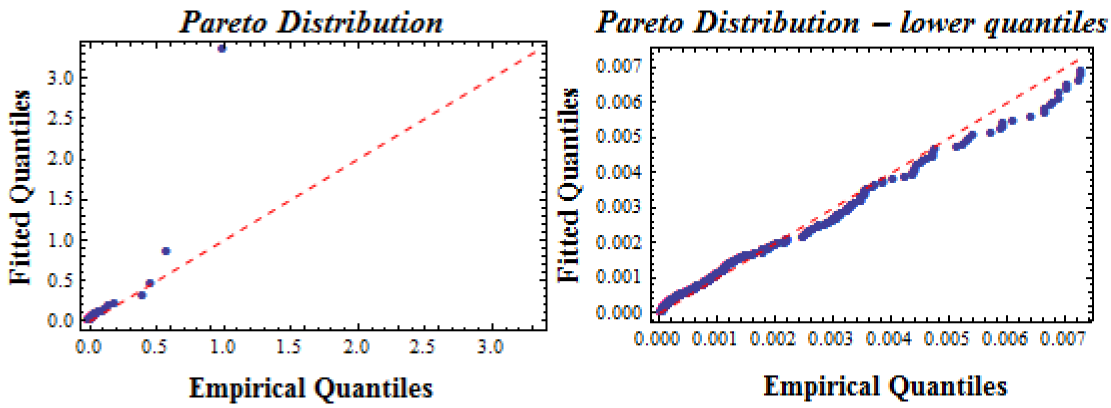
Two-parameter Pareto distribution is perhaps the most notorious heavy-tailed distribution used in financial modeling. KS and AD statistics again reject the null that the data are generated by this distribution. The reason why it is so is conveyed visually by the QQ plots. Most quantiles exhibit almost perfect fit but the Pareto distribution diverges in the tail – it greatly overestimates the rightmost quantiles (the last one is more than three times higher than its empirical counterpart) and that is enough to make the GOFT test statistics extremely high.

The problem signaled by the goodness of fit inspection is visible already in the Table 4.13. Unfortunately the estimate of the shape parameter α is below 1, which means that the fitted distribution has infinite mean. This property renders the distribution useless for further modeling (further details on the issue of infinite mean distributions in operational modeling are reviewed in Embrechts et al. (2006)). Any outcome of the Monte Carlo simulation will be extremely random since losses of potentially infinite magnitude can be drawn from the fitted distribution. We nevertheless run the simulation to give the reader a glimpse on the behavior of the compared risk measures under such circumstances.

Results of the simulation summarized in Table 4.14 match our expectations. Calculated VaR is very high (more than 50% of gross income) while ES totally explodes. Clearly some extremely high aggregate losses have been generated in the tails of the distribution that drive the estimate of ES up to unrealistic heights. But reader should now be aware of the fact that both ES and VaR are useless in this case and that relying upon VaR would be a mistake, even though we can imagine that in some similar cases VaR could seem “reasonable” (e.g. close to 15%). In such cases it might prove useful to have ES estimate because it reveals the infinite mean property rather infallibly[28], as demonstrated in this example (another good idea is to check sensitivity of VaR to small changes in the confidence level, as it tends to be high in infinite mean models).

[28] With a little exaggeration one could say that Eexpected Shortfall’s tendency to explode in such cases can be considered a beneficial property.

Figure 4.9: QQ plots for the fitted Pareto distribution



Source: Author.

Table 4.14: Simulated risk capital charges for Pareto distribution

Pareto Distribution				
α	***ES***	***VaR***	***ES/VaR***	***Equivalent* α**
99.9%	2537.56%	53.71%	4724.81%	93.80%
99%	741.99%	15.68%	4730.59%	42.08%
95%	176.41%	6.90%	2556.43%	13.52%

Source: Author.

4.5.7. The g-and-h distribution

Table 4.15: Estimated parameters for g-and-h distribution

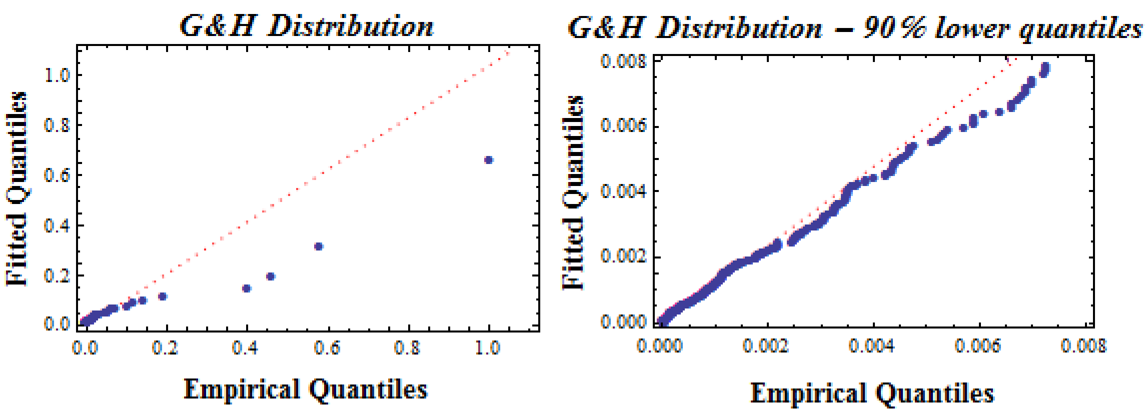
G-and-H Distribution	
$\widehat{A}$	82892.2842
$\widehat{B}$	163481.2403
$\widehat{g}$	2.09245
$\widehat{h}$	0.05479

Source: Author.

We have used the percentile method described in Chapter 2 to estimate the parameters. GOFTs are not calculated due to the fact that the CDF of g-and-h distribution is unknown, as already explained. The QQ plots (as our sole tool for model evaluation in this case) display an acceptable fit and suggest that g-and-h distribution might be suitable for modeling operational risk data. Fit for the 90% lower quantiles very good except for the last quantiles where the QQ plot curves downward below the 45% line. Regarding the tail, several rightmost extreme quantiles are underestimated but at least the distribution follows the trend of the data and does not fail to account completely for the heavy tail, as other distributions did. Judging from the QQ plots g-and-h exhibits a best fit amongst the parametrical distributions we have tried in this section - gamma distribution has fitted the rightmost quantiles better, but it failed badly to model the body, while lognormal distribution represented the best fit for the body of the distribution but severely underestimated the tail.

Implied regulatory capital charge (4.89%) is low but above average of charges computed for other distributions. ES is significantly higher than VaR (by nearly 70%) signalling a heavy tail of the aggregate loss distribution, which means that significantly higher losses lie beyond the VaR threshold. ES statistic evaluates the economic capital charge at 10.35%, which already seems as a reasonable number. Heavier tail is also embodied in the quickly decreasing VaR and ES statistics. These are the characteristics we expect to see in the aggregate loss distribution since they reflect the nature of operational risk data.

Figure 4.10: QQ plots for the fitted g-and-h distribution



Source: Author.

Table 4.16: Simulated risk capital charges for g-and-h distribution

G-and-H Distribution				
α	***ES***	***VaR***	***ES/VaR***	***Equivalent* α**
99.9%	10.35%	4.89%	211.75%	99.57%
99%	3.31%	1.80%	183.51%	95.96%
95%	1.65%	0.98%	168.03%	81.00%

Source: Author.

4.6. Extreme Value Theory distributions

Since the traditional parametric distributions proved to be insufficient to model our operational risk data we have also employed EVT technique in our research. The theory behind EVT has been described in sufficient detail in Chapter II. What is left to explain is our way of estimation of the capital charges for EVT approach. We are using a similar method as in section 4.3 (consult section 2.5 for theoretical background), except for a few adjustments that are made necessary by the fact that EVT does not model whole severity distribution. We have first selected the extreme losses using either Block Maxima Method or Peaks over Threshold Method. Then we fitted the GEV or GPD distribution to the set of extreme losses using probability weighted moments method. To simulate one-year aggregate loss by Monte Carlo simulation we first draw 12 random numbers from fitted Poisson distributions (one for each month) and sum them to N to simulate the frequency. Then we draw N_L random numbers from the empirical distribution function of the data (without the set of extreme observations Z) to simulate the losses from the body of the distribution and N_H random numbers from the fitted GEV or GPD distribution to simulate the extreme events, such that

$$\frac{N_L}{N_H} = \frac{n}{Z},$$

where n is number of observations and Z is number of elements in the set of extreme losses which is dependable on the specific method of selection of extremes[29]. Finally we sum all $N_H + N_L$ simulated losses to obtain the aggregate loss for a single year and repeat the process 50 000 times.

Note that amount of 50 000 simulations wes selected mainly due to technical limitations. Even though this number is typically regarded as sufficient for similar kinds of simulations and in operational risk empirical research one rarely encounters MC simulations exceeding 50 000 – 100 000 years, it can still be said that "the more the better" as increasing number of trials gives a more robust image of an aggregated loss distribution. This holds true especially when one is interested in extremely high quantiles such as 99.9% needed to estimate the regulatory capital charge under Basel II.

[29] This is intuitive since for example if 2% of all losses have been selected as extreme losses, then 2% of simulated losses will come from the fitted distribution and 98% of losses is drawn from empirical sample distribution.

Note that even though empirical sampling works fine for modeling of the losses from the body of the distribution, some fitted parametric distribution could also be used (e.g. Lognormal, Pareto type II distribution or other distribution that would provide a good fit to the dataset without the extreme observations). Nevertheless these losses are more or less inconsequential as they only represent a negligible part of the aggregate yearly loss. With this in mind we decided to use empirical sampling for generation of normal losses in order to decrease computation time.

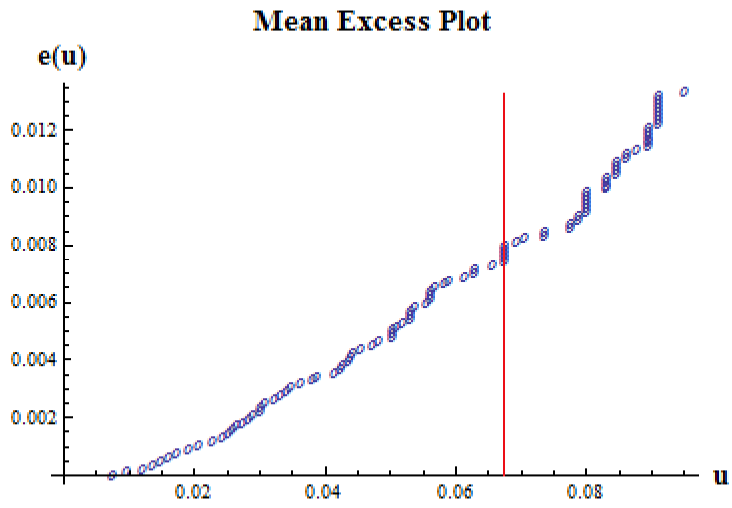
We should definitely note here that there is one potential problem with the EVT approach – the dataset might not be large enough to offer sufficient data points for reliable estimation of GPD or GEV. We could lower the thresholds (or decrease the time block size in case of BMM) to increase sample size of extreme observations and thus lower obtain parameter estimates with lower variance but that would very likely bias the estimation since GPD and GEV are limiting distributions of extreme losses only for sufficiently large thresholds (see Chaudhury (2010) for details on this problem). Since there is a clear trade-off we decided not to use thresholds or methods selecting more than 10% of observations (and even that could too high a percentage).

Few remarks are in order also for the reported goodness of fit statistics – these are incomparable to the ones reported above for the parametric distributions and generally aso incomparable between various EVT methods. The reason is simple – for parametric distributions from previous section whole dataset has been used to estimate the distribution parameters and AD and KS statistics have been calculated using all possible quantiles (one for each observation). But in case of EVT the number of observations used for distribution fitting and for the calculation of test statstics differs method from method (e.g. only 12 observations is used to estimate GEV from under Quarterly Maxima method) hence the definition of test statistics and their critical values change. This is another reason why we are reporting only p-value levels – not to create an impression of comparing the incomparable.

4.6.1. Mean Excess Plot

Figure 4.11 shows the mean excess plot for thresholds up to 10% of the maximum loss amount which corresponds to 98^{th} percentile of the data (again both axes are scaled by the maximum loss). The plot is clearly upward-sloping, suggesting a heavy-tailed distribution and it is approximately linear over whole range of thresholds. Chernobai et al. (2007) proposes a general rule of thumb is to select *u* such that the mean excess function for *x*>*u* is roughly linear, which is not very helpful in this case. Moscadeli (2004) assert that the goal is to detect *u* where the plot is straightening out or, at least, a changes the slope significantly. We highlighted one such *u* with a vertical line, but there is no clear evidence that it is a correct choice and as a reader might have ascertained here, selecting the threshold using mean excess plot is not always possible or at least tends to be somewhat arbitrary.

Figure 4.11: Mean Excess Plot of the data



Source: Author.

Instead of this graphical method we decided to fit more different models and select the most suitable one based on the measures of goodness of fit (in the choice of method we follow Chalupka and Teply (2008)[30]). From this section we can conclude that mean excess plot is insufficient tool for threshold selection and for this purpose analytical methods are desperately needed.

4.6.2. Block Maxima Method – max quarterly losses

First method we have utilized forms the set of extreme observations from the largest losses observed in each quarter. Quarter is perhaps the longest time block we can afford to

[30] But we do not use all of the models presented in this paper, namely we leave out the max 2% losses method (as the results were extremely similar to max quarterly losses method) and method derived from mean excess plot since we do not find its inspection conclusive enough.

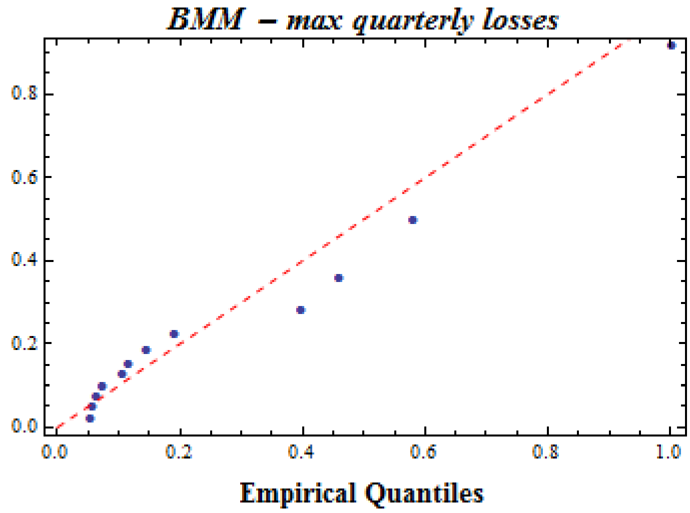
use given the limited time span of our dataset – using this method it shrinks down to 13 observations, which is already a disputable number, with even longer time blocks the number of observations would be by all means too low to estimate GEV reliably.

Table 4.17: Summary of GOFTs and parameter estimates for EVT methods

	Block Maxima Method		**Peaks Over Threshold Method**	
	Max quarterly	*Max monthly*	*Max 5%*	*Max 10%*
KS (p-value)	>0.2	>0.2	>0.1	>0.2
AD (p-value)	>0.2	>0.2	>0.2	>0.2
$\hat{\xi}$ ***(shape)***	0.449	0.777	0.691	0.771
$\hat{\sigma}$	17 170 000	2 488 400	4 755 900	1 999 400
$\hat{\mu}$	17 683 000	1 453 500	4 910 100	2 247 000

Source: Author.

Figure 4.12: QQ plot for fitted GEV distribution (max quarterly losses)



Source: Author.

Table 4.17 summarizes the results of tests of goodness of fit and the estimated parameter values for all EVT methods. In case of maximum quarterly losses method both GOFTs and the QQ plot document a very good match between fitted GEV and empirical quantiles (given the low number of observations and flexibility of GEV this is not so surprising). Estimated shape parameter is positive which indicates heavy tails but lowest amongst EVT methods. The calculated capital charges are the highest so far which confirms our supposition that conventional distributions underestimate the operational losses. VaR is still relatively low (compared to Standardized Approach) but ES is getting close to the results we would normally expect.

Overall, BMM for max quarterly losses seems as a very good model of the aggregate operational losses. Yet, even though it can be considered the best model at this point, one serious drawback is revealed in section 4.9. and that is its over-sensitivity to changes in the underlying dataset, which is implied by the low number of observations used for the estimation of GEV.

Table 4.18: Simulated risk capital charges for BMM (max quarterly losses)

BMM - max quarterly losses				
α	***ES***	***VaR***	***ES/VaR***	***Equivalent* α**
99.9%	10.36%	6.25%	165.76%	99.59%
99%	4.11%	2.67%	153.70%	96.47%
95%	2.37%	1.57%	150.90%	82.51%

Source: Author.

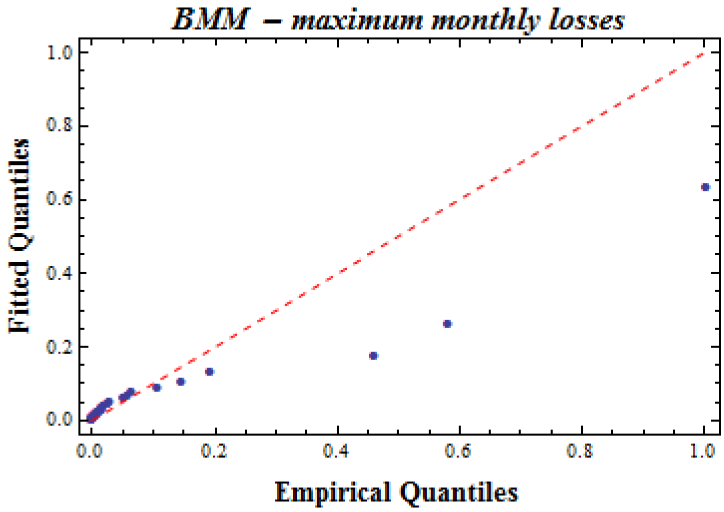
4.6.3. Block Maxima Method – max monthly losses

Second BMM method selects the highest loss from each month to construct the set of extreme losses. Number of observations used is thus 41. Goodness of fit for this method is still perfectly acceptable with p-values of both AD and KS tests above 20%. Nevertheless it is somewhat worse than in case of max quarterly losses method, which is most likely caused by the higher number of observations. Rightmost quantiles of GEV now seem modestly, but systematically, underestimated (one should acknowledge that the quantiles displayed in the Figure 4.13 are by far not as dense as in QQ plots from section 4.5. due to decreased number of observations – in case of this method the steps between quantiles are approximately 2.5% compared to original 0.16% steps).

Estimated capital charges reported in Table 4.19 are quite reasonable and also bring one very interesting comparison. While 99.9% VaR is entirely comparable with the previous BMM method the ES statistic for the same confidence level is now more than two times higher and almost three times as high as VaR. Such is the difference between “the minimum loss in 0.1% worst cases” and “the average loss in 0.1% worst cases”. 7.05% of gross income reported by 99.9% VaR is in fact also an average loss in 0.52% worst cases according to the calculation of Equivalent α.

Obviously, owing to the higher shape parameter (see Table 4.17) this distribution of losses is much more heavy-tailed than the one from max quarterly losses and this feature is clearly indicated by ES. We added the histograms of the two aggregate loss distributions in question to the Appendix (Figure A) to offer a more explicit representation of the discussed phenomena. This result might serve as an example of a situation where VaR is insufficient for distinguishing the riskier of the two different risk profiles.

Figure 4.13: QQ plot for fitted GEV distribution (max monthly losses)



Source: Author.

Table 4.19: Simulated risk capital charges for BMM (max quarterly losses)

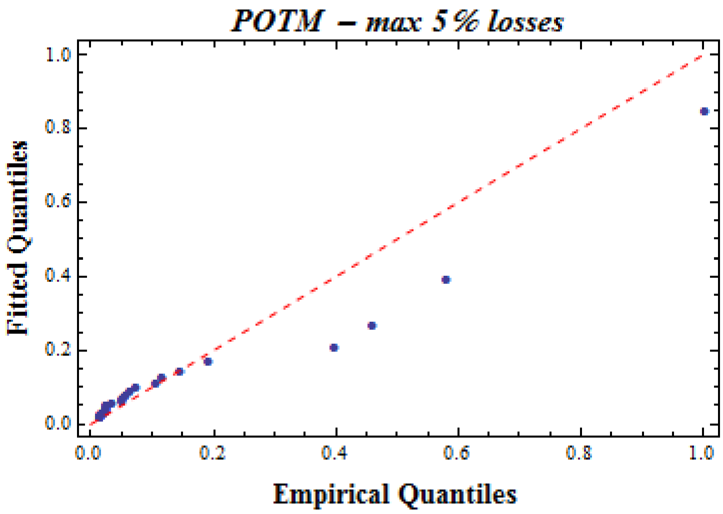
BMM - max monthly losses				
α	***ES***	***VaR***	***ES/VaR***	***Equivalent* α**
99.9%	20.33%	7.05%	288.22%	99.48%
99%	4.60%	1.58%	290.67%	94.64%
95%	1.65%	0.65%	253.00%	71.75%

Source: Author.

4.6.4. Peaks Over Threshold Method – max 5% losses

POTM filters the extreme losses solely by their magnitude, disregarding the order or time of occurrence. In the first method we are modeling the tails by the Generalized Pareto Distribution using highest 5% of data (which amount to 32 observations). Fit of the method to the extreme quantiles of empirical loss distribution is very good, confirming the good overall results of EVT. Compared to BMM the max 5% loss model (and also max 10% loss model) produces significantly higher capital estimates. But the regulatory capital charge still remains in a reasonable range below 15% prescribed by SA, in fact amongst all other results this one seems most consistent with the suppositions of BIS (2010). Yet economic capital charge estimate is already quite high and the gap between the two is very wide, indicating that there is a huge tail risk beyond the VaR level predicted by this model (notice that Equivalent α for 99.9% confidence level is in this case only 99.37%).

Figure 4.14: QQ plot for fitted GPD distribution (max 5% losses)



Source: Author.

Table 4.20: Simulated risk capital charges for POTM (max 5% losses)

POTM – max 5% losses				
α	***ES***	***VaR***	***ES/VaR***	***Equivalent* α**
99.9%	48.78%	13.63%	358.01%	99.37%
99%	10.03%	3.44%	291.73%	94.36%
95%	3.68%	1.56%	235.75%	74.66%

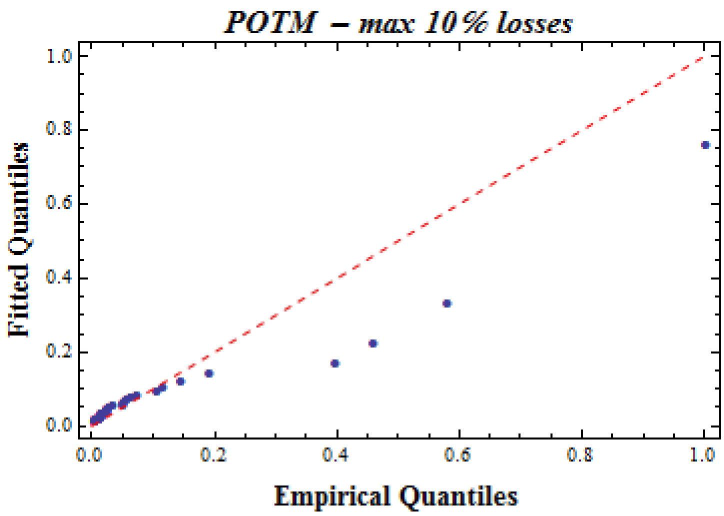
Source: Author.

4.6.5. Peaks Over Threshold Method – max 10% losses

As in BMM case, also for POTM the shape parameter is higher when we use more observations (due to the fact that all of the added observations have a very low magnitude compared to the highest losses, hence the sample mean decreases and variance increases significantly and as a result shape parameter must be higher to model a heavier tail). And also in this instance the fit is little bit worse than for the max 5% losses method. In both cases we prefer – conformably with the GOFTs - the method using less observations as a basic model of operational risk exposure (although in later analysis we will present one disadvantage they have – they are certainly more sensitive to the changes in underlying dataset which turns into a problem when performing scenario analysis and stress testing).

Interestingly also in case of POTM the VaRs (tables 4.20 and 4.21) have been estimated very similarly, in proximity of 15% benchmark. And also in this case the ES statistic is very different (by approximately 16% in terms of gross income) - once again we see the underestimation of risk by VaR, this time for max 5% losses model.

Figure 4.15: QQ plot for fitted GPD distribution (max 5% losses)



Source: Author.

Table 4.21: Simulated risk capital charges for POTM (max 10% losses)

POTM – max 10% losses				
α	***ES***	***VaR***	***ES/VaR***	***Equivalent*** α
99.9%	32.93%	14.65%	224.75%	99.62%
99%	7.72%	2.58%	299.00%	94.54%
95%	2.73%	1.03%	265.68%	73.26%

Source: Author.

4.7. Summary and discussion of LDA results

Table 4.22 constitutes the summary of the main results. We have included in our analysis a wide selection of parametrical distributions that are commonly used for modeling of one-sided risks. First, somewhat obvious, result is the confirmation of the finding of many authors that operational risk data is characterized by heavy-tails and cannot be successfully modeled using the conventional thin-tailed statistical distributions.

Amongst the parametric distributions using the whole sample (non-EVT) the g-and-h distribution was evaluated as the most suitable and potentially very useful for operational risk modeling, even though the regulatory capital charge obtained with this method is very low compared to 15% benchmark from Basel II (recall that beta distribution, despite the more "likable" capital estimate, was ruled out based on goodness of fit measures). In fact it was the only of the examined distributions able to model both body and the tail of the data satisfactorily – most other distributions failed to provide an acceptable fit for the extreme right quantiles even in the case of a good fit in the lower quantiles (namely lognormal distribution due to a much too fast decay in the right tail and the Pareto distribution, which switched to a regime with infinite mean, overestimating the extreme quantiles and rendering the capital estimates useless).

Table 4.22: Summary of estimated regulatory and economic capital charges

Summary of results for α = 99.9%				
Severity distribution	***ES***	***VaR***	***ES/VaR***	***Equivalent α***
Weibull	0.34%	0.33%	103.51%	99.76%
Exponential	1.00%	0.98%	102.45%	99.76%
Lognormal	1.84%	1.25%	147.45%	99.66%
EMPIRICAL	2.66%	2.45%	108.52%	99.73%
Gamma	2.81%	2.55%	110.11%	99.75%
G-and-H	10.35%	4.89%	211.75%	99.57%
Beta	5.31%	5.09%	104.20%	99.75%
BMM - max quarterly losses	10.36%	6.25%	165.76%	99.59%
BMM - max monthly losses	20.33%	7.05%	288.22%	99.48%
POTM - max 5% losses	48.78%	13.63%	358.01%	99.37%
POTM - max 10% losses	32.93%	14.65%	224.75%	99.62%
Pareto	2537.56%	53.71%	4724.81%	93.80%

Source: Author.

Clearly all of the Block Maxima Methods and Peaks over Threshold Methods from Extreme Value Theory provide a better fit for the tails of the data than any of the above-mentioned parametric distributions (indeed the method has been designed specifically for this task so this result only confirms our expectations). Regulatory capital charges under EVT seem to be determined mainly by the type of limiting distribution – for fitted GEV they were 6.25% and 7.05% while for GPD they were slightly below the 15% baseline given by Basel II Standardized Approach. Both of these levels are reasonable and it is not straightforward to judge which method is a better approximation of the true risk exposure.

Economic capital estimates are significantly higher for POTM models, which are clearly more heavy-tailed judging from the ES statistic.

The findings above have one more important implication for the regulatory capital charge (measured as 99.9% Value-at-Risk). By comparing various results and the estimated capital charge from the empirical distribution function as the benchmark we can draw a conclusion that the 15% regulatory capital charge (set as an industry average of capital reserves set against operational losses) is very large compared to the outcomes derived from Loss Distribution Approach. That means that the bank was obliged to keep larger capital buffer against operational losses than what it could try to defend as a reasonable outcome of internal model – a result that confirms the assertion in BIS (2006) that banks will on average be able to decrease their regulatory capital charge by switching from Standardized Approach to Advanced Measurement Approach (the incentives to do so clearly are provided in this case). On the negative side it is likely that the capital allocation was not efficient – since the capital reserve against capital losses was unnecessarily large some of the capital could have been put in better use elsewhere.

An interesting comparison can be made that sheds some further light on the wider problem of fitting the heavy-tailed parametrical distributions to the whole sample. It concerns the results for EVT distributions and Pareto distribution (the similarity between Pareto and Generalized Pareto Distribution is obvious, but their characteristic feature of power-tails is also shared by GEV distribution for $\xi > 0$). We can see that the Pareto distribution has fitted the data relatively well (judging from the QQ plot) but only at a cost of having infinite mean (shape parameter α was than 1). Such situation has never occurred with EVT distributions even though it would be perfectly possible (these distributions have infinite mean for shape parameter ξ greater than 1). Owing to that under EVT approach we could always run a Monte Carlo simulation to obtain meaningful results which was not possible in simple Pareto case, conformably with results of Moscadeli (2004), Jobst (2007) or Chalupka and Teply (2008). The ability to model the tail with decreased risk of running into an infinite mean model is in fact part of the rationale behind the idea of EVT and limiting the sample to the extreme observations.

Crucial part of the results concerns the Expected Shortfall as our representation of the class of coherent risk measures. Following points summarize the main conclusions:

1) We have demonstrated that ES produces reasonable outcomes and that it is by all means an applicable tool for operational risk modeling (except for infinite mean models, as described in point 3). It has some theoretical properties superior to VaR (it is subadditive in the sense of Artzner et al. (1999) and utilizes more information from the underlying loss distribution) and moreover it shares some of VaR's attractive features – it is an intuitive and easily comprehensible risk measure, it is suited for the description of extreme quantiles of loss distribution and it is easy to implement. Moreover it is straightforward to calculate the statistics such as "equivalent α" that make direct comparisons with VaR possible. Considering the ease of implementation it provides additional information to a risk modeler with little or no extra computational effort.

2) Difference between VaR and ES is larger the heavier the tails of the loss distribution is. ES is by definition always larger than VaR of the same confidence level (except for special case when VaR is equal to maximum loss), we have used two different tools to quantify the difference between the two risk measures. First is a simple ratio ES/VaR and second is an "Equivalent α", i.e. a confidence level α such that for some $\overline{\alpha}$ we have $VaR_{\overline{\alpha}}(X) = ES_{\alpha}(X)$. We have shown that the degree of difference between ES and VaR ranges from irrelevant (e.g. exponential or Weibull distributions) to extremely significant (g-and-h or EVT models), mainly depending on the heaviness of the tail of simulated distribution of aggregate operational losses.

3) We have shown that ES gives useful information on the tail structure of underlying loss distribution and thus provide additional insight into the modeled risk profile. It is a particularly suitable tool for evaluation of the heavy-tailedness (which is, in a sense, equivalent to riskiness) of the loss distribution where VaR fails to be one. A great illustration for this point is the difference between estimated loss distributions of the two BMM methods. Models seem almost equally good based on the goodness of fit measures and also the reported VaRs are nearly the same (reporting 6.25% and 7.05% capital charge). But there is a crucial difference in the tail behavior of the two loss distributions that is unaccounted for by VaR but captured by ES well, reporting 10.36% and 20.33% capital charges, hence evidently

indicating that the loss distribution derived from max monthly method is associated with much riskier position.

4) In the infinite mean models ES tends to diverge to extremely high values as exposed on the case of Pareto distribution model. In fact, from the formulation of ES it is clear that it is not even defined for such models and hence cannot be used to calculate the operational risk exposure in these cases. But we have explained why also VaR produces very unreliable and potentially misleading results in such cases (e.g. Embrechts et al. (2006) say that infinite mean models should be completely avoided in operational risk modeling). We pointed out that ES might serve as a first indication of the danger of a bad (diverging) model - whereas VaR might conceal the infinite mean property (e.g. in cases when all extreme losses lie behind the VaR threshold) ES statistic will typically be driven to extraordinarily high values by the most severe losses.

5) ES can be used either as a complement or as an alternative[31] to VaR in economic and regulatory risk capital estimation.

   - Rationale for using ES as a complement is that it compensates the main shortcoming of VaR – and that is its ignorance of the observations that lie behind the threshold for some arbitrarily chosen α (one could say that VaR ignores the most severe 100*(1-α)%* of losses). Besides the model selection, as demonstrated in section 4.5, it could also be useful for evaluation of risk positions e.g. when updating the model for new observations and assumptions, as ES is a more suitable tool for the model comparison than VaR.

   - Using ES as an alternative to VaR provides additional theoretical benefits of coherence[32] as discussed in chapter 3. The subadditivity property might be crucial from practical viewpoint especially when calculating the total risk exposure as the sum of capital charges estimated for individual business line, as required by Basel II. Yet for these purposes the adjustment (downwards) of the confidence level α may be necessary, otherwise the calculated capital charges might be prohibitively large. In fact, the extremely high confidence level of

[31] One could remark here that VaR must be calculated in any case since it is a prerequisite for the calculation of Expected Shortfall. When we speak about ES as an alternative to VaR, we have in mind using it instead of VaR for the calculation of regulatory or economic capital charge.

[32] Concept of coherence in the sense of Artzner et al. (1999) is explained in chapter 3.

99.9% for Basel II regulatory capital charge is set to rectify the above-mentioned drawbacks of VaR. The idea is to leave as few observations behind the VaR threshold as possible (since larger α implies less significant information loss). Because ES accounts even for most extreme losses smaller confidence levels are admissible. One possible way is to set the confidence level close to estimated "Equivalent α" so that the ES capital charges are at least comparable to the charges calculated with 99.9% VaR[33] – such method would provide the benefits of ES while not ruining the consistency of Advanced Measurement Approach under Basel II.

4.8. Stress Testing

This section of the thesis is directed to compare the sensitivity of Value-at-Risk and Expected Shortfall (as a representation of coherent risk measures) to changes in the underlying dataset of operational losses. Our methodological approach is very similar to the scenario analysis stress testing method as described in Jorion (2007) or Rippel (2008).

Relevance of this exercise stems from Basel regulatory requirements, in BIS (2010a) we can read that the AMA of a bank requires the use of four data elements which are: internal loss data; external data; scenario analysis and business environment and internal control factors. At this point we are focusing on the third element, specifically on the question how is the use of custom scenarios reflected in the risk measures under scrutiny. Scenarios might include broad range of events from any operational risk category, such as inappropriate or failed business practices, terrorist attacks, rogue trading, electricity failures or unavailability of the interbank clearing and payment system (Rippel (2008)). Nevertheless it is not within a scope of this thesis to provide an in-depth scenario analysis and it would not even be possible without disclosing the details on the BANK's business activities, economic background and operational management practices. Scenarios must be carefully designed to reflect such specifics of the individual company's environment. Our interest goes only so far to see how different risk measures under different loss distribution assumptions react to the changes in the underlying dataset. Following Rippel

[33] Lowest equivalent α for $VaR_{99.9\%}$ amongst our simulations has been evaluated to 99.37% for POTM (max 5% losses) method.

and Teply (2008) we will define several hypothetical custom scenarios, i.e. operational losses that will be mixed with the internal dataset.

All of the scenarios we are considering are based on historical operational risk events, so even though they have not happened to the BANK itself their reoccurrence (or occurrence of similar situation with similar loss magnitude) cannot be ruled out. Nevertheless the main criterion for the choice of events was the variety of their magnitudes rather than variety of event types. In fact since this is not a typical scenarios analysis we do not need to concern ourselves with the relevance of given scenarios that much – we only want to stress test the selected models and compare the outcomes for the scrutinized risk measures and for this goal it suffices to build a list of scenarios with different loss orders.

Table 4.23: Table of custom losses used for stress testing

Custom Losses			
ID	***Event type***	***Magnitude***	***Event description***
x	*x*	*0.58%*	*(Maximum loss in the original dataset)*
1	Process management failure	0.72%	Software loss due to incorrect rounding (KB)
2	External fraud	2.10%	Theft – Procházka
3	Internal fraud	3.31%	Unauthorized Trading - Kweku Adoboli (UBS)
4	Internal fraud	11.09%	Unauthorized Trading - Jerome Kerviel (SG)

Source: Author.

Table 4.24: List of stress tests based on custom losses

Stress Tests	
Test number	***Added observations***
Test 1	Loss 1
Test 2	Loss 2
Test 3	Loss 3
Test 4	Loss 4
Test 5	Loss 1, Loss 2, Loss 3, Loss 4

Source: Author.

Table 4.23 summarizes the hypothetical events; three of them coincide with Rippel (2008). Two are based on cases of unauthorized trading activity, one on the case of Jerome Kerviel from early 2008 (Société Général) and second on the more recent case of Kweku

Adoboli in UBS[34]. Third loss is associated with the theft of cash in amount of approximately 560 million CZK committed in Czech Republic by Frantisek Procházka (several banks have been affected, e.g. CSOB has suffered a loss exceeding 200 million CZK). Least severe loss in the table is derived from the event from another Czech commercial bank Komerční Banka which has accumulated a huge loss over several years of using an incorrect setup of rounding of interbank transaction fees.

To get an idea on the severity of the hypothetical losses they are stated in proportion to the gross income ratio so that they can be directly compared to the aggregate losses calculated by Monte Carlo simulation approach. Note that despite the fact that our hypothetical losses are all based on historical operational risk events their magnitude indicated in Table 4.23 might not match their original magnitude. That is because we have rescaled the losses to fit the size of the BANK based on the ratio of the total assets of a bank where the event happened to total assets of the BANK, hence each loss was calculated as

$$Loss\ amount = Original\ loss\ amount * \frac{BANK's\ total\ capital}{Total\ capital\ of\ the\ affected\ bank}.$$

Using the four custom losses we have performed five stress tests. First test adds all four events as additional observations into the dataset (hence the total number of observations increases to 637) whereas other tests always add only a single observation as shown in Table 4.24. For each test we have re-estimated the parameters of selected distributions and performed a new simulation of aggregate loss distribution to calculate VaR and ES at 99.9% confidence level. This way we can observe how sensitive the capital charge estimate is to the changes in the underlying dataset. Notice though that the new observations introduced into the dataset are all higher than the maximum loss in the original dataset. They vary in terms of magnitude, none the less they are all extreme events.

The list of distributions used for this part of research reflects the results from the previous part. We include the three best-performing models – g-and-h, BMM (max quarterly losses) and POTM (max 5% losses). Because the timing of the losses matters in case of Block Maxima Method, we needed to use a randomizing algorithm to determine

[34] For some basic information on this event reader might consult internet sources like http://edition.cnn.com/2011/BUSINESS/09/15/switzerland.bank.lost/index.html. For detailed description of other three events see Rippel (2008) and references therein.

the exact time of occurrence for each loss. Additionally we have appended a lognormal distribution to the list. This distribution had a solid fit in the body but underestimated the extreme quantiles - we want to use it as a demonstration of the behavior of the two risk measures under light-tailed[35] distribution.

4.9. Summary and discussion of stress tests results

Table 4.25: Results of stress tests for lognormal distribution

Lognormal Distribution, α=99.9%				
Test number	***ES***	***VaR***	***ES/VaR***	***Equivalent α***
Original	1.84%	1.25%	147.45%	99.66%
Test 1	1.71%	1.29%	131.83%	99.69%
Test 2	1.97%	1.36%	145.08%	99.65%
Test 3	2.03%	1.41%	144.06%	99.65%
Test 4	2.06%	1.56%	131.93%	99.72%
Test 5	3.38%	2.28%	148.10%	99.67%

Source: Author.

We begin with the model that reveals only very week sensitivity to the tested scenarios. Even though there is a tractable and consistent increasing trend in VaR and ES the addition of extreme losses to the dataset does not generate a very significant effect in these statistics. Test 5 which includes all four new extreme events at once does not even double the estimated capital charges (recall that Loss 4 from Table 4.23 is nearly 20 times higher than maximum loss from the original dataset) and neither does it make the tails heavier as seen in ES/VaR and Equivalent α statistics. Notice how surprisingly stable these comparisons of ES and VaR are – for all five models they remain very close to their original values from section 4.6. Obviously, performing scenario analysis with such model does not make much sense since the distribution still puts most weight on the main body of observations and extreme losses only cause minor deviations. It is not flexible enough to even provide a reasonable model of the heavy-tail, let alone to analyze the changes in its structure.

[35] Lognormal distribution is sometimes classified as "heavy-tailed", here we use the term light-tailed purely in a sense that it grossly underestimated the extreme right quantiles for the application at hand (i.e. it did not prove as heavy-tailed enough to model out data).

Table 4.26: Results of stress tests for g-and-h distribution

G-and-H Distribution, α=99.9%						
Test number	$\hat{g}$	$\hat{h}$	***ES***	***VaR***	***ES/VaR***	***Equivalent α***
Original	2.07	0.05	10.35%	4.89%	211.75%	99.57%
Test 1	2.08	0.06	9.90%	5.20%	190.48%	99.58%
Test 2	2.12	0.09	19.96%	9.96%	200.44%	99.62%
Test 3	2.16	0.10	27.09%	10.24%	264.61%	99.45%
Test 4	2.14	0.24	149.22%	33.63%	443.72%	99.25%
Test 5	2.14	0.32	230.19%	87.54%	262.96%	99.47%

Source: Author.

Table 4.25 reviews the results for the g-and-h distribution. Test 1 which adds an observation of slightly higher magnitude as the maximum loss in the dataset has a result perfectly comparable with the original result (even though the parameter estimates are both slightly higher, the Monte Carlo simulation produced a loss distribution with a little lighter tail). Test 2 and Test 3 give a similar estimate of regulatory capital but the latter case is significantly riskier with ES statistic higher by almost 10% of gross income. Inclusion of extreme losses in tests 4 and 5 leads to very intense capital estimates, both regulatory and economic. Unsurprisingly the decay in the tail of the loss distribution is faster in test 5 (as reflected in ES/VaR and equivalent α statistics) where several high-magnitude losses are added then in case 4 where the added loss creates an "outlier" that is very far from the rest of the observations. In this case we see relatively stable estimates of Equivalent α for the first two tests, then it begins to decrease as the structure of the tail becomes heavier.

Table 4.27: Results of stress tests for BMM (max quarterly losses)

BMM (max quarterly losses), α=99.9%					
Test number	$\hat{\xi}$	***ES***	***VaR***	***ES/VaR***	***Equivalent α***
Original	0.45	9.36%	6.25%	149.62%	99.69%
Test 1	0.48	19.61%	10.72%	182.97%	99.60%
Test 2	0.73	109.76%	38.75%	283.23%	99.48%
Test 3	0.81	197.52%	70.38%	280.66%	99.49%
Test 4	0.94	1747.73%	200.95%	869.75%	98.90%
Test 5	0.61	2012.27%	223.46%	900.51%	98.87%

Source: Author.

Results for Block Maxima Method are less satisfactory. One weakness of the model has been exposed and that is its extreme sensitivity to the changes in the dataset. Test 1 nearly doubles both economic and regulatory capital charges while remaining tests already imply unrealistic economic capital estimates. This caveat stems naturally from the fact that this method only uses 13 observations from the whole dataset to find the parameters of GEV distribution, which means that any change in the set of extremes causes a significant alteration of the model. Quickly increasing heaviness of the tail can be observed both in the growing estimates of shape parameter ξ and in the declining trend in equivalent α, signaling a widening gap between the VaR measure and the most extreme observations in the simulated loss distribution. This behavior of the model has unpleasant implications for practical application – most importantly this model does not seem fit for purposes of scenario analysis which should be an organic part of operational risk modeling according to Basel II rules (this conclusion has also been reported by Rippel (2008)). Secondly such model can prove too volatile with the changes in the internal dataset (e.g. with arrival of new observations, since operational risk modeling is a time-continuous process and the models need to be regularly updated) implying an undesirable state of unpredictability of capital allocation needs. If lognormal was an example of distribution too insensitive in the tails, then this BMM specification can serve as an example of over-sensitive model.

Table 4.28: Results of stress tests for POTM (max 5% losses)

POTM (max 5% losses), α=99.9%					
Test number	$\hat{\xi}$	***ES***	***VaR***	***ES/VaR***	***Equivalent α***
Original	0.69	48.78%	13.63%	358.01%	99.37%
Test 1	0.66	52.13%	13.79%	378.03%	99.32%
Test 2	0.81	63.99%	17.64%	362.78%	99.35%
Test 3	0.86	75.10%	20.45%	367.24%	99.47%
Test 4	0.95	390.24%	43.55%	896.07%	98.87%
Test 5	0.80	798.67%	60.85%	1312.59%	98.40%

Source: Author.

Also in case of POTM model we can observe an increasing shape parameter ξ as the GPD distribution is trying to accommodate the new observations by the heavier tail (for the Test 5 the lower value of ξ is compensated by higher values of scale and location parameters, but the provided fit is rather poor in this case). But unlike in previous case, this time also the estimates of ES and VaR seem reasonable. In the first two tests the regulatory capital remains very similar to the original model while the ES slightly increases reflecting a fatter tail. Economic capital estimates for tests 4 and 5 produced by POTM method are significantly larger than in g-and-h case, despite that regulatory capital estimates are comparable (as discussed above, BMM method estimates are unreasonably high). The difference increases as we add larger losses to the dataset. That could mean one of two following – either the g-and-h method underestimates the extreme quantiles (then POTM would give a more realistic picture of the true risk exposure under) or the POTM method suffer from the same problem as BMM, only to a lesser extent. We are inclined to believe that the first supposition is true. Disregarding the last two tests the results for POTM actually seem quite consistent and reasonable and the method also provides an acceptable fit to all the modified datasets whereas the QQ plots for g-and-h distribution reveal underestimation[36] of far-right quantiles similarly as in section 4.5 (even a more distinct one with more extreme losses added to the dataset).

Based on these considerations we favor POTM as a most suitable method for modeling of our dataset of operational losses. For this method the ES exhibits similar sensitivity to the changes in underlying dataset as VaR except for the scenario 4 and 5 with an extremely severe event (approximately than 20 times higher than a maximum loss in the original dataset) added as new observation - in that case ES increases much more than VaR as the tail of the simulated aggregate loss distribution becomes extremely heavy. Hence even though parts of our results confirm the assertion of some authors (e.g. Yamai and Yoshiba (2002) or Broda and Paolella (2009)) who say that ES is generally somewhat less stable than VaR, such property is most pronounced when catastrophic losses come into the picture and thus it would cause troubles mainly in the context of stress testing and scenario analysis. Yet the worse convergence properties might be problematic also for regular modeling, especially when the underlying severity distribution is strongly heavy-tailed. Following our results this problem was only tractable for the stress tests thus we do not consider it a fatal drawback for practical application of ES, but some further research on this issue would definitely be helpful.

4.10. Final remark

We close our research by returning to the beginning – to our hypotheses as stated in Introduction part. First hypothesis H_0^A concerned the use of EVT in operational risk modeling. We have shown that some of the EVT models have produced best results amongst the statistical techniques used in this thesis. It proved suitable for approximation

[36] QQ plots for g-and-h and POTM can be found in Appendix.

of extreme quantiles of loss distributions which is paramount to operational risk data modeling and the POTM method also exhibited an ability to provide consistent capital estimates under our stress tests. We can hence confirm the findings of Moscadelli (2004), Jobst (2007), Chalupka and Teply (2008) or Dahen et al. (2010), that EVT provides a reasonable and consistent framework for quantitative operational risk management, which also clearly confirms our first hypotheses.

More importantly, regarding the second hypotheses H_0^B (on the applicability of coherent risk measures in operational risk modeling), we have successfully employed Expected Shortfall as a tool for consistent estimation of risk capital requirements against operational losses. In empirical operational risk research ES has not been used much yet but there already are some examples of its successful application such as Giacometti et al. (2008) or Lee and Fang (2010) – unfortunately they only use ES for capital charge reporting and lack a theoretical treatment with deeper comparison to VaR. We have taken the discussion on Expected Shortfall further and demonstrated the benefits of ES compared to VaR in context of operational risk. Even though we also pointed out some caveats we still conclude that it is an extremely valuable tool for quantification of operational risk exposure and thus we can answer our second hypothesis affirmatively.

4.11. Suggestions for further research

Our work can be extended in several ways. First, some of its features of this thesis have been determined by the technical limitations that may eventually prove surmountable. This concerns mainly the Monte Carlo simulations of aggregate operational losses. In case of EVT we have simulated 50 000 years, which might not be enough for some very heavy-tailed severity distribution to produce sufficient amounts of extreme events. But primarily, in ideal case we would like to run the simulation itself at least several thousand times in order to obtain more robust estimates of ES and VaR along their standard errors. That would moreover allow us to report the risk measures in confidence interval instead of having just their point estimates. Such information is critically needed to compare the estimation error and convergence properties of ES and VaR, nevertheless it is extremely demanding computationally, especially in case of EVT models where a single simulation can run several hours. Even though the comparative analysis of this kind between ES and VaR has been done by several authors (e.g. Yamai and Yoshiba (2002)) it has never been performed in the specific context of operational risk modeling and typically has only been done using the simulated data and conventional statistical distributions - performing such simulations for EVT or g-and-h distributions is a good lead to further research on VaR and coherent risk measures.

Another point of interest is the out of sample performance of ES compared to VaR. As VaR estimates are generally regarded as statistically more stable (see e.g. Sarykalin et al. (2008)) ES might fall short compared to VaR in this respect. We would definitely be interested to assess practical relevance of this argument for operational risk management in some application to real data (our dataset was unfortunately too small to answer such questions).

Further, an extension to concept of coherent risk measures can be found in "distortion risk measures" (basic introduction can be found in Bellini and Caperdoni (2007) or Dowd and Blake (2006)) – these quantitative tools calculate the risk exposure as a mean value of a distorted (by some non-decreasing distortion function g: $[0,1] \rightarrow [0,1]$) survival function $g\left(\bar{F}(X)\right)$ of the aggregate loss distribution X. Distortion risk measure can be proven to be coherent if its distortion function satisfies certain properties. Even though they do not focus specifically on the extreme quantiles of loss distribution they can prove useful as complementary measures as they may provide some additional valuable information on the risk exposure. That is because contrary to VaR and ES, some of the distortion risk measures sustain the stochastic dominance of second and higher orders (see Hurliman (2004) for explanation and some practical examples). We encourage other researchers to explore the potential of these risk measures for the operational risk modeler's toolkit.

Conclusions

In our research we have addressed several issues related to operational risk modeling. Our two main hypotheses have been formulated as follows:

H_0^A: *Extreme Value Theory provides a suitable framework for quantitative operational risk management and produces reasonable and consistent models of operational risk exposure.*

H_0^B: *Expected Shortfall (as a representation of coherent risk measures) is an applicable and consistent tool for estimation of operational risk capital charges under variety of statistical modeling methods and can be used as a suitable complement or alternative to Value-at-Risk.*

Our dataset of operational losses reflects all the stylized facts on operational risk data - it is strongly leptokurtic, right-skewed, characterized by heavy tails and dominated by few high-magnitude observations. We fitted a list of statistical distributions to the data and similarly to de Fontnouvelle et al. (2005), Cruz (2002) or Moscadelli (2004) we observed very poor results. None of the parametric distributions using the whole data sample was able to provide satisfactory fit to the tail of the empirical distribution (even not the heavy-tailed distributions like lognormal, Weibull or Pareto).

One exception is a generalized 4-parameter g-and-h distribution which provided a solid overall fit to the whole dataset (hence we can confirm its applicability along with Jobst (2007), Rippel (2008) or de Fontnouvelle (2005)). But contrary to Rippel (2008) who found it consistent also under scenario analysis we concluded on the basis of stress tests that it is not flexible enough for consistent operational risk modeling as its fit got progressively worse with inclusion of more extreme losses in the dataset (resulting in underestimation of the operational risk exposure).

We evaluated Extreme Value Theory as the most suitable method of operational risk modeling. We should stress that the choice of the correct model is very data-dependent and thus such conclusions cannot be generalized – yet good results for this method in the context of operational risk have also been reported by many other researchers including Moscadelli (2004), de Fontnouvelle (2003) or Chalupka and Teply (2008). Max 5% losses (Peaks over Threshold) method provided best overall results (including the stress tests) amongst all employed models. We can conclude the first hypothesis with an affirmative answer.

Regarding hypothesis B, we have first given the rationale on the use of a coherent statistic of Expected Shortfall in operational risk modeling. We have identified the caveats of VaR and shown that they are closely related to the lack of sub-additivity and implied non-coherence of VaR. One grave practical problem with VaR is its ignorance of the worst case events in the loss distribution and resulting incapability of evaluation of tail risk. ES has much better theoretical properties and represents a natural answer to the most serious shortcomings of VaR. It makes use of more information from the underlying loss distribution and hence provides a better insight into the analyzed risk profile.

Overall ES produced reasonable capital estimates and was extremely useful for inter-model comparison. We have seen examples of pairs of loss distributions that seemed similar in terms of VaR but that represented entirely different risk profiles due to the differences in the heaviness of their tails. In these cases ES proved as an invaluable tool for distinction between different levels of tail risk and hence in this regard it can be considered superior to VaR.

We have however exposed some drawbacks of ES too. This statistic tends to be more volatile than VaR under stress tests (we have scrutinized the stability of the two risk measures for several models by adding hypothetical extreme losses to the original dataset) which is in line with results of Yoshiba and Yamai (2002). Since ES uses even the most extreme losses from the simulated distribution it can also be expected that its estimates have larger variance and slightly worse convergence properties than those of VaR (especially under extremely heavy-tiled distributions). One question also arises with reporting of ES for regulatory purposes – it is not straightforward how to choose an appropriate confidence level α for ES calculation, but 99.9% suggested might be often associated with prohibitively large capital charge estimates (we suggested that α can be possibly determined such that ES for this α is approximately equal to 99.9% VaR).

Even though ES is not flawless and some further research is still needed to clarify certain issues, we still strongly believe that Expected Shortfall is a suitable tool for estimation of economic operational risk capital (more so than VaR), but also can potentially be used for evaluation of regulatory capital. We have produced rich evidence on the applicability and versatility of ES and the results indicate that in certain cases it is a clear improvement over VaR. The basic conclusion is hence quite apparent – ES, if not an alternative, is a priceless complement to VaR in quantitative operational risk management.

In a sense we have illustrated the thought of Guegan and Tarrant (2011) who also claim that multiple different risk measures are needed to properly analyze the risk associated with the tails of the loss distribution. Whereas they recommend a regulatory rule of reporting five different risk measures, we are more modest in our conclusions and propose that Value-at-Risk can be complemented by a single coherent risk measure, Expected Shortfall, for a more reasonable determination of both regulatory and economic capital requirements.

REFERENCES

- Acerbi, C., Nordio, C., and Sirtori, C.(2001). "*Expected shortfall as a tool for financial risk management*," Working paper, http://www.gloriamundi.org/var/wps.html.

- Acerbi, C., and Tasche, D. (2001). "*Expected shortfall: A natural coherent alternative to value at risk*," Economic Notes 31 (2-2002), 1–10, http://www.gloriamundi.org/var/wps.html.

- Acerbi, C. and Tasche, D. (2002). "*On the Coherence of Expected Shortfall*," Journal of Banking and Finance, 2002, 26 (7), 1487–1503.

- Albanese, C. (1997). "*Credit Exposure, Diversification Risk and Coherent VaR*," Working paper, Department of Mathematics, University of Toronto.

- Artzner, P., Delbaen, F., Eber, J., and Heath, D. (1997). "*Thinking Coherently,*" RISK 10, pp. 68-71.

- Artzner, P., Delbaen, F., Eber, J., and Heath, D. (1999). "*Coherent Measures of Risk*," Mathematical Finance 9, pp. 203-228.

- Aue, F., and Kalkbrener, M. (2006). "*LDA at work: Deutsche Bank's approach to quantifying operational risk*," The Journal of Operational Risk 1/4, 49–93.

- Bellini F., and Caperdoni, C. (2007). "*Coherent distortion risk measures and higher-order stochastic dominances*," North American Actuarial Journal, 11(2):35.

- Biagini, F. and Ulmer, S. (2009). "*Asymptotics for operational risk quantified with expected shortfall,*" ASTIN Bulletin 39, 735–752.

- BIS (1999). "*A New Capital Adequacy Framework*," www.bis.org.

- BIS (2001a). "*The New Basel Capital Accord,*" www.bis.org.

- BIS (2001b). "*Operational Risk,*" Consultative document to Basel II capital accord, www.bis.org.

- BIS (2006). ''*International Convergence of Capital Measurement and Capital Standards,* www.bis.org.

- BIS (2010a). "*Operational Risk – Supervisory Guidelines for the Advanced Measurement Approaches*," www.bis.org.

- BIS (2010b). "*Sound Practices for the Management and Supervision of Operational Risk*," www.bis.org.

- Broda, S. A. and Paolella, M. S. (2009). "*Calculating Expected Shortfall for Distributions in Finance*," Mimeo, Swiss Banking Institute, University of Zurich.

- Burnecki, K., Misiorek, A. and Weron, R. (2010). "*Loss Distributions*," MPRA Paper No. 22163.

- Chaudhury, M. (2010). "*A review of the key issues in operational risk capital modeling,*" The Journal of Operational Risk, 5/3, 37-66.

- Cizek, P., Hardle, W., and Weron, R., eds. (2011). "*Statistical Tools for Finance and Insurance,"* 2nd edition, Springer, Heidelberg.

- Chernobai, A., Fabozzi, F., and Rachev, S. (2007). "*Operational Risk: A Guide to Basel II Capital Requirements, Models, and Analysis,*" John Wiley & Sons, New Jersey.

- Chernobai, A., Menn, C., Rachev, S., and Truck, S. (2005). "*Estimation of Operational Value-at-Risk in the Presence of Minimum Collection Thresholds*," Technical report, University of California at Santa Barbara.

- Chalupka, R., Teplý, P. (2008). "*Operational Risk Management and Implications for Bank's Economic Capital – A Case Study*," IES Working Paper 17/2008, IES FSV, Charles University.

- Cruz, M. G. (2002). "*Modeling, Measuring and Hedging Operational Risk*," John Wiley & Sons, New York.

- Chorafas, D. (2006). "*Economic Capital Allocation with Basel II*," Elsevier, Oxford.

- Dahen, H., Dionne, G., and Zajdenweber, D. (2010). "*A practical application of extreme value theorzý to operational risk in banks,*" The Journal of Operational Risk, 5/2, 63-78.

- de Fontnouvelle, P., Deleus-Rueff, V., Jordan, J. and Rosengren, E. (2003). "*Using loss data to quantify operational risk,*" Federal Reserve Bank of Boston, Working Paper.

- de Fontnouvelle, P., Jordan, J., and Rosengren, E. (2005). "*Implications of Alternative Operational RiskModeling Techniques*," Technical report, Federal Reserve Bank of Boston and Fitch Risk.

- Degen, M., Embrechts, P., and Lambrigger, D. D. (2007). "*The quantitative modeling of operational risk: between g-and-h and EVT,"* Astin Bulletin 37, 265–291.

- Delbaen, F. (2002). "*Coherent Risk Measures on General Probability Spaces*," In Essays in Honour of Dieter Sondermann, pp. 1–37. Berlin: Springer.

- Dhane, J.L.M, Gooaverts, J., and Kaas R. (2003). "*Economic Capital Allocation Derived From Risk Measures,*" North American Actuarial Journal, 7(2):44–56.

- Dhaene, J., Laeven, R.J.A., Vanduffel, S., Darkiewicz, G., and Gooaverts, M.J. (2006). "*Can a Coherent Risk Measure be Too Subadditive?*" Journal of Risk and Insurance 75, 365–386.

- Dowd, K., and Blake, D. (2006). "*After VaR: The Theory, Estimation, and Insurance Applications of Quantile Based Risk Measures,*" Journal of Risk and Insurance, 73(2): 193-229.

- Dutta, K., and Perry, J. (2007). "*A Tale of Tails: An Empirical Analysis of Loss Distribution Models For Estimating Operational Risk Capital*," Federal Reserve Bank of Boston, Working Papers No. 06-13.

- Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). "*Modelling Extremal Events for Insurance and Finance,*" Springer, New York.

- Embrechts, P., Furrer, H., and Kaufmann, R. (2003). "*Quantifying regulatory capital for operational risk*," Derivatives Use, Trading & Regulation 9(3), 217/233.

- Embrechts, P., Chavez-Demoulin, V., and Neslehova, J. (2006). "*Infinite Mean Models and LDA for Operational Risk,*" Journal of Operatinal Risk, 1, 3-25.

- Embrechts P, Neslehova J, Wuthrich M.V. (2009). "*Additivity properties for Value-at-Risk under Archimedean dependence and heavy-tailedness,*" Insurance: Mathematics and Economics, 44, 164-169.

- Giacometti, R., Rachev, S., Chernobai, A., and Bertocchi, M. (2008). "*Aggregation issues in operational risk*," The Journal of Operational Risk 3/3, 3–24.

- Gourieroux C., and Jasiak J. (2010). "*Value at Risk,*" Working Paper, York University, http://dept.econ.yorku.ca/~jasiakj/papers/ait.pdf.

- Goodheart C. (2001). "*Operational Risk,*" Special Paper 131, Financial Markets Group, London: London School of Economics.

- Guegan, D., and Tarrant, W. (2011). "*On the Necessity of Five Risk Measures*," Quantitative Finance Papers, arXiv.org.

- Hoaglin, D. (1985). "*Summarizing Shape Numerically: The g-and-h Distributions*," in Hoaglin D., Mosteller, F. and Tukey, J., eds., *Exploring Data Tables, Trends, and Shapes*, pp.461/513 , John Wiley & Sons, Inc., New York.

- Heyde, C., Kou, S. and Peng, X. (2007). "*What Is a Good Risk Measure: Bridging the Gaps between Data, Coherent Risk Measures, and Insurance Risk Measures*," Preprint, Columbia University.

- Hurliman, W. (2003). "*Conditional Value-at-Risk Bounds For Compound Poisson Risks and a Normal Approaximation,*" Journal of Applied Mathematics, 2003/3, 141–153.

- Hurliman, W. (2004). "*Distortion Risk Measures and Economic Capital*," North American Actuarial Journal 8, 86-95.

- Inui, K., and Kijima, M. (2005). "*On the significance of expected shortfall as a coherent risk measure,*" Journal of Banking & Finance, 29, Risk Measurement, pp. 853-864.

- Jobst, A. A. (2007). "*Operational Risk — The Sting is Still in the Tail but the Poison Depends on the Dose*," IMF Working paper 07/239, International Monetary Fund.

- Jorion, P. (2007). "*Value-at-Risk: The New Benchmark for Managing Financial Risks*," 3rd ed., McGraw-Hill, New York.

- Krause, A. (2003). "*Exploring the limitations of value at risk – how good is it in practice?*", Journal of Risk Finance 4, 19–28.

- Krokhmal, P., Palmquist, J., and Uryasev, S. (2001). "*Portfolio optimization with conditional value-at-risk objective and constraints*," Journal of Risk 4(2), pp.43-68.

- Lee, W. and Fang, C. (2010). "*The measurement of capital for operational risk in Taiwanese commercial banks,*" The Journal of Operational Risk, 5/2, 79–102.

- Mejstrik, M., Pecena, M. and Teply, P. (2008). "*Basic Principles of Banking*," Karolinum Press, Prague.

- Moscadelli, M. (2004). "*The modeling of operational risk: experience with the analysis of the data collected by the Basel committee,*" Technical Report 517, Banca d'Italia.

- Peters, G.W., Sisson, S.A. (2006). "*Bayesian inference, Monte Carlo sampling and operational risk.*" The Journal of Operational Risk 1/3, 27-50.

- Pflug, G. (2000). "*Some remarks on the value-at-risk and the conditional value-at-risk*," in Uryasev, S. (2000), *Probabilistic Constrained Optimization: Methodology and Applications,* Kluwer Academic Publishers. http://www.gloriamundi.org/var/pub.html.

- Power M. (2005). "*The Invention of Operational Risk*," Review of International Political Economy 12(4): 577–599.

- Rippel, M., Teplý, P. (2008). "*Operational Risk - Scenario Analysis,*" IES Working Paper 15/2008, IES FSV, Charles University.

- Rippel, M. (2008). "*Operational Risk Scenario Analysis*," Diploma Thesis, IES FSV, Charles University.

- RiskMetrics (2001). "*Return to RiskMetrics: The Evolution of a Standard*, " http:/www.riskmetrics.com.

- Rockafellar, R., and Uryasev, S. (2000). "*Optimization of Conditional Value-at-Risk*," Journal of Risk 2, 21–42.

- Rockafellar, R., and Uryasev, S. (2002). "*Conditional value-at-risk for general loss distributions*," Journal of Banking and Finance 26, 1443–1471.

- Sarykalin S., Serraino G., Uryasev S. (2008) "*VaR vs CVaR in risk management and optimization*," INFORMS tutorial.

- Schevchenko, P. and M. Wuthrich (2006). "*The structural modelling of operational risk via Bayesian inference: Combining loss data with expert opinions*," CSIRO Technical Report Series, CMIS Call Number 2371.

- Tukey, J. (1977). "*Exploratory Data Analysis,*" Addison-Wesley, Reading.

- Wang, S.S. (2002) "*A Universal Framework For Pricing Financial and Insurance Risks*," ASTIN Bulletin, 32, 213-234.

- Wirch, J.L., Hardy, M.R. (2000). "*Ordering of risk measures for capital adequacy*," Institute of Insurance and Pension Research, University of Waterloo, Research Report 00-03.

- Yamai Y., and Yoshiba, T. (2002a). "*Comparative Analysis of Expected Shortfall and Value-at-Risk: Their Estimation, Decomposition, and Optimization*," Monetary and Economic Studies, January, pp. 87-121.

- Yamai Y., and Yoshiba, T. (2002b). "*On the Validity of Value-at-Risk: Comparative Analysis with Expected Shortfall*," Monetary and Economic Studies, January, pp. 57-85.

- Yamai Y., and Yoshiba, T. (2002c). "*Comparative Analyses of Expected Shortfall and Value-at-Risk (2): Expected Utility Maximization and Tail Risk*," Monetary and Economic Studies, January, pp. 95-115.

- Yamai Y., and Yoshiba, T. (2002d). "*Comparative Analyses of Expected Shortfall and Value-at-Risk: Their Validity Under Market Stress*," Monetary and Economic Studies, January, pp. 181-237.

PUBLIC INFORMATION SOURCES

- www.edition.cnn.com (CNN website)

- www.bis.org (Bank for International Settlements website)

SOFTWARE USED FOR CALCULATIONS[37]

- **Easy Fit**
- **Gretl**
- **Mathematica**
- **Microsoft Excel**
- **Microsoft Visual Basic for Applications**
- **R**

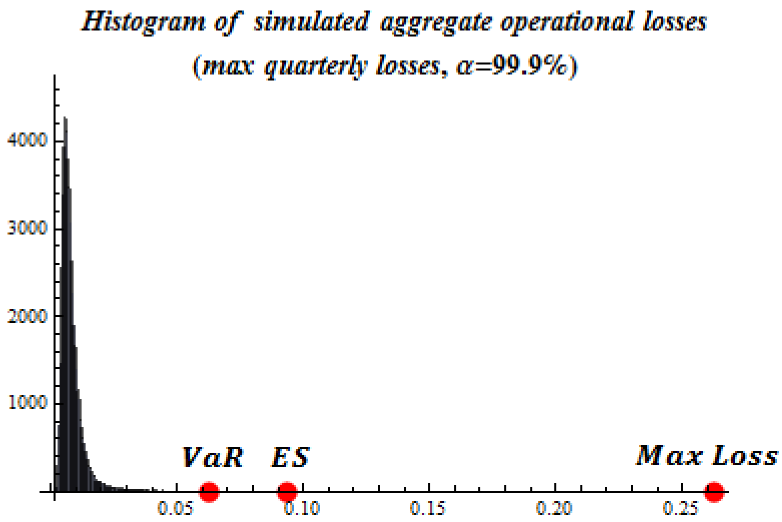
[37] The software names are listed in alphabetical order. They are the registered trademarks of registered owners.

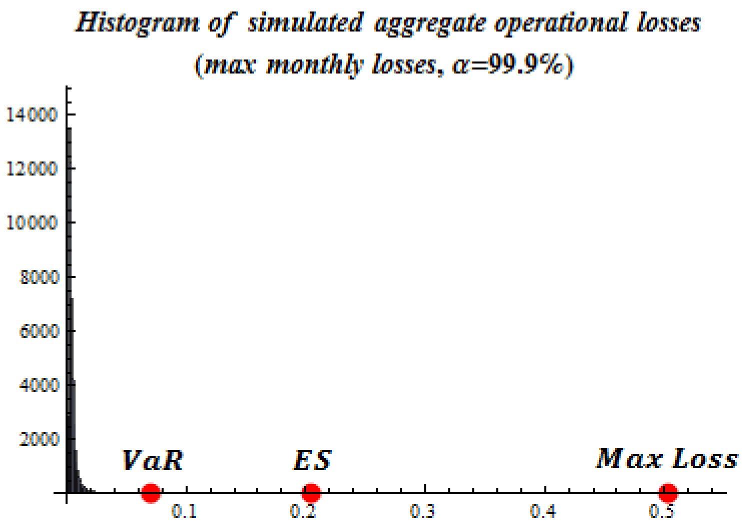
LIST OF ABBREVIATIONS

AD test	Anderson-Darling test
AMA	Advanced Measurement Approach
BIA	Basic Indicator Approach
BMM	Block Maxima Method
CVaR	Conditional Value-at-risk
ES	Expected Shortfall
ESM	Empirical Sampling Method
EVT	Extreme value theory
GEV	Generalized Extreme Value (distribution)
GOFT	Goodness of Fit Tests
GPD	Generalized Pareto Distribution
KS test	Kolmogorov-Smirnov test
LDA	Loss Distribution Approach
MLE	Maximum Likelihood Estimator
MTL	Median Tail Loss
PDF	Probability density function
POTM	Peaks Over Threshold Method
PWM	Probability-weighted moments
QQ plot	Quantile-quantile plot
SA	Standardized Approach
VaR	Value-at-Risk

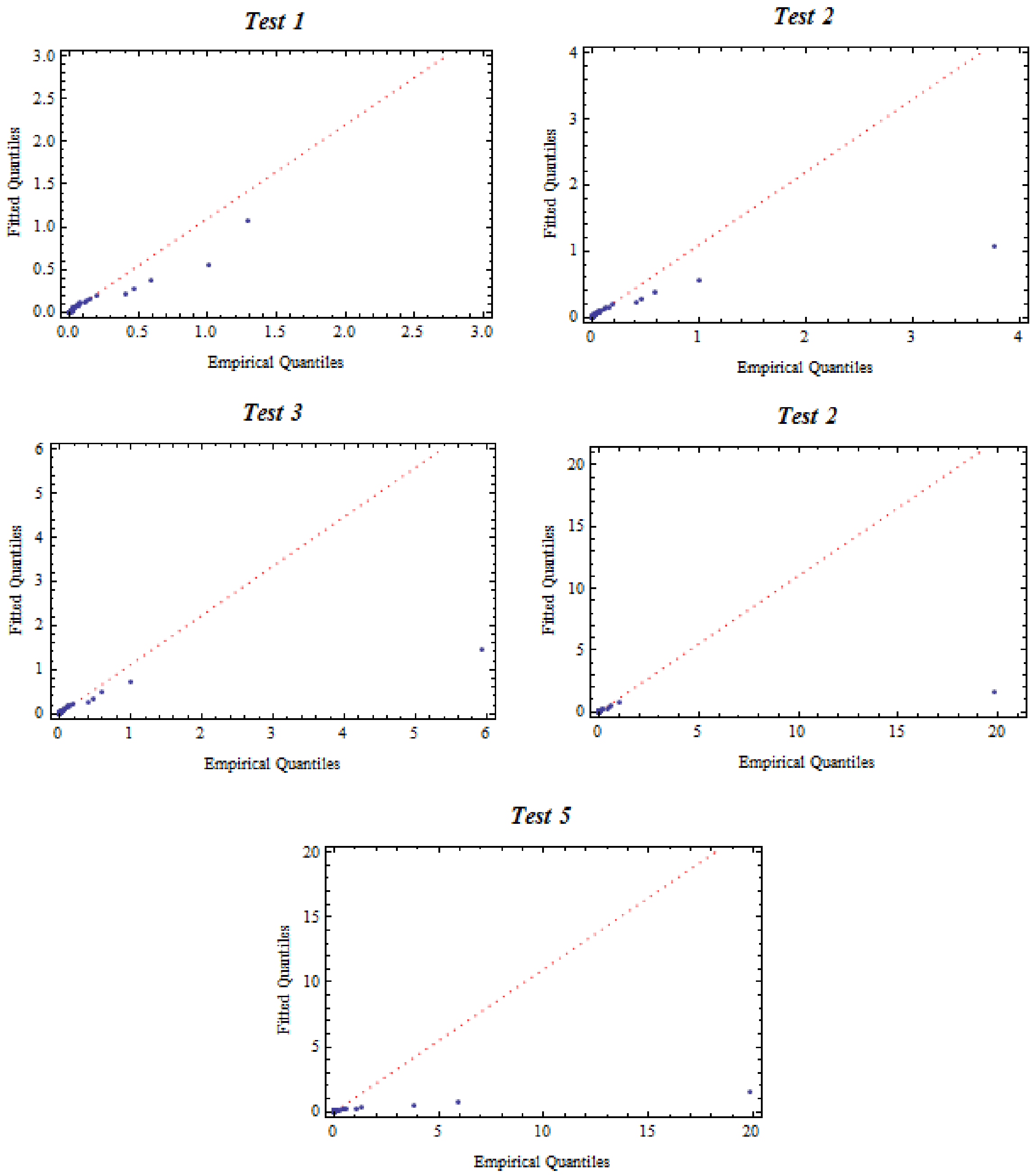
APPENDIX

APPENDIX A: The upper and lower panels display the histograms of simulated 1-year aggregate operational losses derived from BMM max quarterly losses and max monthly losses models respectively. Loss severity is scaled by the bank's gross income ratio as defined in Basel II rules (see Chapter 1 for details).VaR, ES and maximum simulated loss are highlighted in the figures. Even though the reported VaRs are nearly the same, heavier tail of the second distribution is manifest in the larger dispersion of extreme losses (largest 50 out of 50000 simulated for 99.9% confidence level) behind the VaR threshold and it is clearly reflected in the ES statistic.





APPENDIX B: The QQ plots of fitted g-and-h distributions for all five tests from Section 4.8 are displayed in this figure. One can easily observe that the worsening fit for extreme quantiles as more extreme losses are added to the dataset, which suggest limited flexibility of g-and-h distribution for modeling catastrophic losses. This phenomenon leads to underestimation of the capital charge derived from aggregate loss distribution in test 2-5.



APPENDIX C: The QQ plots for GPD distribution (Peaks over Threshold Method with max 5% losses) estimated under tests 1-5 are shown below. Underestimation of right-most quantile is visible as in g-and-h case, but it is much less pronounced. Better fit of this method on the catastrophic losses is possible owing to the fact that it only models the tail of the distribution (specifically the highest 5% of observations) while omitting the main body of the data in the estimation. As a result the simulated aggregate loss distribution is more representative of the true operational risk exposure.