LUND UNIVERSITY

FACULTY OF SCIENCE

DEPARTMENT OF MATHEMATICAL STATISTICS

MASTER THESIS

On Modeling Operational Risk Using Extreme Value Theory

Author:
Valentin KÄSTNER

Supervisor:
Nader TAJVIDI

June 2, 2016

Abstract

The main goal of this thesis is to show how operational risk can be measured if even the use of standard extreme value theory fails to explain single catastrophic events in the tail of the distribution. Against the background of regulatory requirements imposed by the Basel Accords, an Advanced Measurement Approach (AMA) is developed for a dataset of operational losses occurred in US businesses between 1985 and 2008.

Two alternative approaches are described for modeling the loss frequency when the losses are reported by the month. A copula approach is applied to capture dependence among different loss distributions corresponding to different event types. The resulting 99.9% Value-at-Risk, which determines the capital requirement, is compared to a model assuming perfect dependence.

Keywords: Operational Risk, Risk Management, AMA, Extreme Value Theory, Peaks over Thresholds, Value at Risk, Copula, Outlier, Markov chains, SMA, Basel II

Populärvetenskaplig sammanfattning

Operativ risk syfta till risken av ekonomiska förluster som uppstår genom felaktiga interna processer, mänskliga brister, systemstörningar eller genom externa händelser. Finansiella institutioner som banker eller försäkringsbolag ska enligt lag indentifiera, mäta, styra och övervaka sådana risker för att förutse risken av en betydande ekonomisk förlust. Dessutom måste de finansiella institutionerna inneha eget kapital vars belopp beror på risknivån.

I denna masteruppsats används statistiska metoder som extremvärdesteori och copulas för att utveckla en så kallad Advanced Measurement Approach (AMA) modell för att kvantifiera risknivån. Särskilt beskrivs hur den vanliga extremvärdesteorin kan anpassas så att extrema avvikande värde kan analyseras på ett lämpligt sätt inom teorin.

Contents

Acknowledgements

I would like to express my deep gratitude to my supervisor, Dr Nader Tajvidi, for his valuable advice and useful critiques as well as providing me with operational loss data. I would also like to thank Dr Ainura Tursunalieva from Swinburne University of Technology and Professor Paramsothy Silvapulle from Monash University for their valuable support regarding the data material. Finally, I wish to thank my family, friends and girlfriend for their encouragement throughout my study.

Chapter 1

Introduction

Every business faces a number of different risks. A farmer struggles with bad harvest and a sharp drop in selling prices, manufacturing companies experience losses from machine break downs and banks are exposed to credit defaults. Despite the differences of these risks, there is a certain type of risk which threatens every business in one way or another. This type of risk is called operational risk and, roughly speaking, describes the risk of losses resulting from human misbehaviour and natural processes, such as the damage due to natural disasters.

Unlike the farmer and the manufacturing company, banks are subject to legal control when it comes to risk management. This is a result of their outstanding role in creating liquidity in economy. By forcing financial institutions to hold a reasonable amount of equity capital, the regulators aim at reducing the risk of a financial collapse as recently experienced in 2007-2008. The exact capital charge depends on the degree of risk, whose calculation often requires the use of sophisticated statistical methods.

The goal of this thesis is to examine how the common model for operational risk can be improved to obtain a better fit and more reliable capital charge estimates. Two datasets are investigated to demonstrate when the regular Peaks-over-Threshold approach fails to model the tail events.

The structure of this thesis is as follows. In Chapter 2 the role of risk management in view of the capital requirements imposed by the Basel Accords are discussed. Chapter 3 gives a deeper insight into the regulatory and modeling aspects of operational risk. Before a Loss Distribution Approach is applied to the data in Chapter 5, Chapter 4 aims at introducing the theoretical instruments needed for modeling operational risk losses. The modeling results are then discussed in Chapter 6 followed by a short conclusion.

Chapter 2

Risk Management

The role of risk management in financial institutions has increased significantly during the past decades. Events such as the recent subprime crisis, the dot-com bubble in 2000 or the 2008 Société Générale trading loss led to financial turmoil, which over and over arose the question how these dramatic incidents can be avoided. In this chapter a short introduction into the process of risk management is given, mainly being based on the excellent explanations in McNeil et al. [2005] and Hull [2015] as well as on own practical experiences within the field of risk management. The main tasks of risk management are discussed, important types of risk are presented and the regulatory framework initiated by the Basel Committee on Banking Supervision is outlined.

> "Risk management is a discipline for living with the possibility that future events may cause harm." – Kloman [1992, p. 305].

This definition gives a good explanation what risk management is and what it is not. The expression *living with* underlines the permanent presence of risk that cannot be changed by any risk management department, no matter how sophisticated the used methods are. Risk management means accepting risk and knowing how it can be dealt with. It does not mean, risk shall be avoided at any cost. Particularly, as taking risks often comes with the chance of making profit.

The risk management process involves different types of activities. First, it is necessary to identify all types of risks the institution is exposed to. Next, these risks are assessed, e.g., by considering historical data in order to find a statistical model which describes current risk. The validity of these models have to be monitored on a regular basis and need to be reported to executive directors. Risks are controlled and mitigated according to

the bank's risk appetite. For a detailed description of basic principles of risk management in the context of operational risk, we refer to Basel Committee on Banking Supervision [2011].

2.1 Financial Risk

Financial institutions face various types of risk. This statement is not limited to banks and insurances, but includes all sorts of business companies. In general, financial risks can be divided into four big categories.

Credit Risk

If a borrower fails to pay predefined repayments (he "defaults"), the lender suffers a financial loss. The possibility of such an event is called credit default risk or counterparty risk. Although it is common practice to use default risk as a synonym for credit risk, the latter is defined more broadly. It also contains the risk of changes in the credit quality (downgrade risk) and in the credit spread (spread risk). Credit risk is generally seen as the most important risk category for banks.

Market Risk

Market risk is the possibility of suffering financial losses from changes of market prices. It includes the change of stock prices or stock indices (equity risk), of interest rates (interest rate risk) and currency rates (currency risk). Additionally, some companies, in particular within the energy industry, face high risks from oil price movements. This type of risk can be classified as commodity risk and is another subcategory of market risk.

Operational Risk

Fraud, natural disasters or disruption of business can produce severe losses for any company. In particularly serious circumstances, it may even make them fail. As an example, the disastrous Fukushima tsunami in 2011 can be brought forward, which caused the nuclear power plant operator Tokyo Electric Power Company (TEPCO) to be saved by the government. These manifold types of risks are summarised as operational risk. It is discussed more throughly in Chapter 3.

Liquidity Risk

Losses which are caused by the lack of liquidity are a consequence of liquidity risk. This is the case if some party wants to trade an asset, but no other market participant is willing to trade for this asset. McNeil et al. [2005, p. 3] compares liquidity with "oxygen for a healthy market". Most of the time, it is inconspicuous. However, if it is lacking, it might entail dramatic consequences.

2.2 The Basel Accords

Although it is said that modern banking started in early Renaissance, the history of banking regulation does not start before the 20th century in many countries. To take but one example, Germany introduced banking regulation in the 1930s after suffering severe losses from the World Economic Crisis in 1929.[1] The necessity of a common international regulation framework arose when the markets became increasingly international. As a consequence of a number of financial disruptions such as the abolition of the Bretton-Woods system of fixed exchange rates in the early 1970s, the Basel Committee on Banking Supervision (BCBS) was founded by the central bank governors of the G10 countries in 1974.[2] The motivation for establishing the BCBS "was and is to enhance financial stability by improving supervisory knowhow and the quality of banking supervision worldwide." (Basel Committee on Banking Supervision [2015, p. 1]) It is important to note that decisions made by the BCBS do not have legal force, but can be seen as guidelines and recommendations. The actual implementation into national laws is conducted by national authorities.

The first important step towards financial stability came in 1988 when the Basel Capital Accord (widely referred to as Basel I) was released. It asked banks for ensuring a minimum capital ratio of capital to risk-weighted assets of 8%. However, this approach was soon considered to be crude and risk-insensitive and needed to be revised.

The 1996 amendment to Basel I and the introduction of Basel II in 2006 remedied this problem by allowing banks to use internal models, which are based on the Value-at-Risk, a highly controversial risk measure that attracted a lot of attention in the 1990s but had already been in use decades before its peak of popularity.[3] The second Basel

[1] See Bundesanstalt für Finanzdienstleistungsaufsicht [2016].

[2] The Group of Ten (G10) consists of eleven industrialised countries that discuss financial matters on a regular basis. Today the BCBS members come from 27 different countries, including e.g. Sweden, Germany and the United States.

[3] More details about the history of the VaR can be found in Holton [2002].

Accord is considered as a revolutionary step towards today's risk management. Not only has it improved the risk sensitivity, but it also extended the risk management process itself by introducing a three pillar concept consisting of *minimum capital charge*, *supervisory review process* and *market discipline*. In particular, the release of Basel II is of highest importance as far as the subject of this thesis is concerned, since Basel II was the first regulatory publication to address the consideration of operational risk. More precisely, it contains the details of the Advanced Measurement Approach that is looked at in detail throughout this work.

As a consequence of the financial crisis in 2007-08 and the decrease of capital requirements due to the use of internal models, another revision of the regulatory standards became necessary. The third regulatory framework in Basel Committee on Banking Supervision [2010] (Basel III) was published in December 2010, but its implementation is still ongoing. Basel III tightens banks' capital requirements, but the general approaches for measuring operational risk remain unaffected. A fourth revised version of the Basel standard is expected to follow in the foreseeable future.

Chapter 3

Operational Risk

Operational risk (OpRisk) is often wrongly defined as the risk of loss neither belonging to credit risk nor market risk. The correct definition in the Basel II framework, see Basel Committee on Banking Supervision [2006, p. 144], is as follows:

Definition 1 (Operational Risk)**.** *The risk of loss resulting from inadequate or failed internal processes, people and systems or from external events. This definition includes legal risk, but excludes strategic and reputational risk.*

A popular example for economical losses due to OpRisk are losses resulting from the September 11, 2001, terror attacks. According to Hull [2015, p. 479], the Bank of New York lost 140 million USD because of the damage the attacks entailed.

Compared to other financial risks, operational risks are more difficult to identify and to control. While in market risk it is possible to adjust the risk profile by buying and selling risky and conservative assets, OpRisk can not be controlled in an equally simple manner. By nature, all business companies are exposed to operational risks and a reduction of this risk to zero requires a reduction of all business activities.

Nevertheless, there are certain ways for a bank to reduce its OpRisk. Examples include a domicile move to a geographically less risky location, the establishment of a general awareness of operational risks within the company, or the improvement of error-prone processes. These actions have in common to be tedious and complex, such that the implementation would be very time-consuming and the immediate effects of the change might be hard to measure. In particular, it is unclear how these changes enter the capital charge calculations.

Basel II provides three different approaches to calculate the required capital. They vary greatly in their risk-sensitivity and complexity.

3.1 The Basic Indicator Approach

The simplest method presented in Basel II is called the Basic Indicator Approach (BIA). The average of all positive gross incomes of the past three years GI_1, GI_2, GI_3 is multiplied by a constant factor α. Mathematically, this can be expressed by

$$K_{BIA} = \frac{1}{n} \sum_{i=1}^{3} \alpha \cdot \max(GI_i, 0),$$

where $n = card\{i \in \{1, 2, 3\} : GI_i > 0\}$ is the number of positive gross incomes in the past three years. The value of α is currently set to 15% by the Basel Committee.

3.2 The Standardised Approach

A slightly more advanced way of computing the capital charge is obtained by using the Standardised Approach (TSA). In similarity with the BIA the Gross Income serves as a proxy for the scale of business operations. But in contrast to the BIA, which uses the same factor for all business lines, the TSA allows for differences, i.e. the gross income needs to be calculated for each business line individually and is then multiplied with a business line specific factor β. Table 3.1 lists the beta factors which are currently prescribed by the Basel Committee. The corresponding mathematical formulation is

$$K_{TSA} = \frac{1}{3} \sum_{i=1}^{3} \max \left(\sum_{j=1}^{8} G_{i,j} \beta_j, 0 \right).$$

Table 3.1: Beta Factors in TSA

Business Lines	**Beta Factors**
Corporate Finance (β_1)	18%
Trading & Sales (β_2)	18%
Retail Banking (β_3)	12%
Commercial Banking (β_4)	15%
Payment & Settlement (β_5)	18%
Agency Services (β_6)	15%
Asset Management (β_7)	12%
Retail Brokerage (β_8)	12%

3.3 The Advanced Measurement Approach

The third and most sophisticated method is called the Advanced Measurement Approach (AMA). It allows the financial institutions to model operational risk more individually and does neither ask for a specific model nor for a specific proceeding. This is in contrast to, e.g., credit risk, where a number of pre-specified parameters (probability of default, loss given default and exposure at default) is estimated. Naturally, the tolerance of models that can be used, is limited and the bank needs to prove the model consistency to the regulatory authority.

In theory there are three alternative ways of dealing with an AMA. Since the Internal Measurement Approach (IMA) and the Scorecard Approach are rarely used, the popular Loss Distribution Approach (LDA) is discussed exclusively. The LDA separates the problem of finding the distribution of the sum of all N operational losses $X_1, \ldots, X_N$ that occur in the next 12 months

$$L = \sum_{i=1}^{N} X_i$$

into two subproblems. The 12-month loss frequency N and the loss severity X_i are assumed to be independent and are modeled individually. These two models can be tied together with the aid of Monte Carlo simulation and the Value-at-Risk at level 99.9% is estimated by the 99.9%-quantile of the Monte Carlo sample. The capital charge is defined to be

$$K_{AMA} = VaR(99.9\%) - EL, \tag{3.1}$$

where EL refers to the expected loss from OpRisk within a 12-months horizon. Figure 3.1 illustrates this relationship.

In comparison to the BIA and TSA, the AMA allows for a more proper subdivision by introducing the concept of event types, see Table 3.2. Losses are classified according to the Business line/Event type matrix and the entries are called units. The idea is to model each unit separately and to aggregate all unit-specific capital requirements to one total number, i.e. the total minimum capital requirement. In addition to modeling historical loss data accurately, banks must satisfy a number of additional requirements to qualify for the AMA. These requirements involve the use of external data, a scenario analysis by experienced risk management experts and the capture of internal control factors to take the current bank's operational risk profile into account.

Despite these additional demands, the large degree of freedom in the modeling pro-

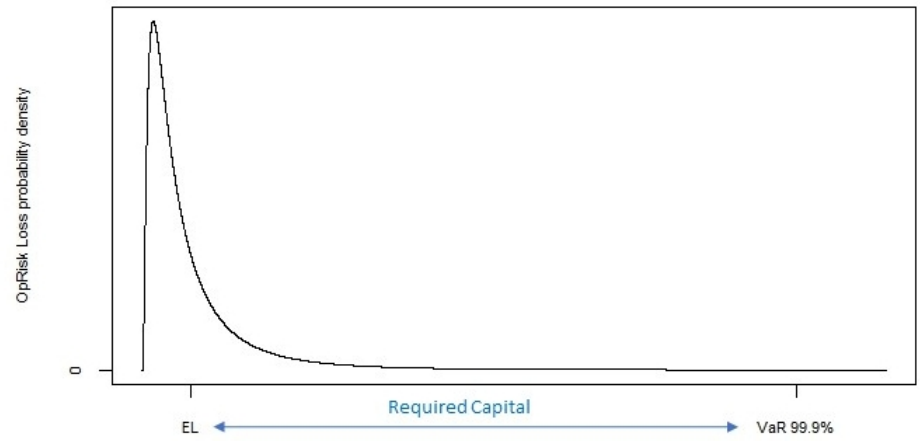

Figure 3.1: Relationship between expected loss, Value-at-Risk and capital charge

cedure made the AMA occupy a special place in literature on OpRisk. The early 2000's saw a vast number of publications proposing various approaches to deal with the AMA. Influential papers are, amongst others, Moscadelli [2004], Aue and Kalkbrenner [2006] and Frachot et al. [2004], with the two latter publications describing the loss distribution approach from a practitioner's point of view for Deutsche Bank and Crédit Lyonnais, respectively. A compilation of interesting articles can be found in Cruz [2004], which also contains a copula approach in Di Clemente and Romano [2004], used to model the dependence amongst different modeling units. Extensive books addressing the AMA and a wide range of theoretical tools are, e.g., Cruz [2002], McNeil et al. [2005] and Bolancé et al. [2012].

By allowing the financial institutions to experiment freely with different models, the Basel Committee hoped for a speedy convergence towards a sophisticated standard model. Unfortunately, despite the extensive coverage of this topic in research papers during the past two decades, these expectations have not been met yet, roughly ten years after the implementation of Basel II. As a consequence, recent discussions argue for an abolition of the AMA:

> "A recent review of the measures related to banks' operational risk modelling practices and capital outcomes revealed that the Committee's expectations failed to materialise. Supervisory experience with the AMA has been mixed. The inherent complexity of the AMA and the lack of comparability arising from a wide range of internal modelling practices have exacerbated variability in risk- weighted asset calculations, and have eroded confidence in risk-weighted

> capital ratios. The Committee has therefore determined that the withdrawal of internal modelling approaches for operational risk regulatory capital from the Basel Framework is warranted." – Basel Committee on Banking Supervision [2016, p. 1]

Table 3.2: Loss event types classification as defined by the Basel Committee on Banking Supervision [2006, Annex 9]

Event-Type Category	**Definition**
Internal fraud	Losses due to acts of a type intended to defraud, misappropriate property or circumvent regulations, the law or company policy, excluding diversity/discrimination events, which involves at least one internal party
External fraud	Losses due to acts of a type intended to defraud, misappropriate property or circumvent the law, by a third party
Employment Practices and Workplace Safety	Losses arising from acts inconsistent with employment, health or safety laws or agreements, from payment of personal injury claims, or from diversity / discrimination events
Clients, Products & Business Practices	Losses arising from an unintentional or negligent failure to meet a professional obligation to specific clients (including fiduciary and suitability requirements), or from the nature or design of a product
Damage to Physical Assets	Losses arising from loss or or damage to physical assets from natural disaster or other events
Business disruption and system failures	Losses arising from disruption of business or system failures
Execution, Delivery & Process Management	Losses from failed transaction processing or process management, from relations with trade counterparties and vendors

3.4 The Standardised Measurement Approach (SMA)

As a result of inadequate and often unreliable models in the AMA, the Basel Committee addressed the Standardised Measurement Approach (SMA) in March 2016. The aim of the SMA is to introduce a simple standardised measure that is both risk sensitive and easily comparible among different banks. The discussion about the SMA is still on a consulting level, such that it is likely to find a slightly modified version in the final regulatory guideline. We give a brief review of the latest proposal presented in Basel Committee on Banking Supervision [2016].

Similar to the BIA and TSA the SMA will be based on a financial statement proxy, the so-called Business Indicator (BI), which shall reflect the bank's exposure to operational risk. The BI resembles the Gross Income, but they are not the same. It is built as the sum of three components, namely the *Interest, Lease and Dividend Component* (ILDC), the *Services Component* (SC) and the *Financial Component* (FC), which are all averaged over the past three years:

$$\text{BI} = \text{ILDC}_{Avg} + \text{SC}_{Avg} + \text{FC}_{Avg}.$$

The mathematical definitions of these three components involve a number of accounting ratios and confusing formulas. For the sake of clarity, we omit these definitions and refer the interested reader to Basel Committee on Banking Supervision [2016, p. 5].

The SMA uses the concept of five different buckets, which are determined by the size of the BI as described in equation (3.2). The minimum regulatory capital is calculated as

$$\text{SMA Capital} = \begin{cases} \text{BIC}, & \text{if Bucket 1} \\ 110m + (\text{BIC} - 110m) \cdot \underbrace{\log\left(\exp(1) - 1 + \dfrac{\text{LC}}{\text{BIC}}\right)}_{\text{Internal Loss Multiplier}}, & \text{if Buckets 2-5,} \end{cases}$$

where BIC and LC are the Business Indicator Component and Loss Component, respec-

tively, and the abbreviation m stands for million. The BIC is defined as

$$\text{BIC} = \begin{cases} 0.11BI, & \text{if } BI \leq €1bn \text{ (Bucket 1)} \\ 110m + 0.15(BI - 1bn) & \text{if } €1bn < BI \leq €3bn \text{ (Bucket 2)} \\ 410m + 0.19(BI - 1bn) & \text{if } €3bn < BI \leq €10bn \text{ (Bucket 3)} \\ 1.74bn + 0.23(BI - 1bn) & \text{if } €10bn < BI \leq €30bn \text{ (Bucket 4)} \\ 6.34bn + 0.29(BI - 1bn) & \text{if } €30bn < BI \text{ (Bucket 5)}, \end{cases} \tag{3.2}$$

i.e., the BIC is a continuous function of the BI with progressive increase of the marginal impact of the BI. The function equations were found by analysis. The loss component is obtained from

$$\text{LC} = 7 \cdot \text{ATAL} + 7 \cdot \text{ATAL}_{10} + 5 \cdot \text{ATAL}_{100}$$

with ATAL being the average total annual loss. The specifications ATAL_{10} and ATAL_{100} denote the average total annual losses calculated from only including loss events above €10 million and €100 million, respectively. Hence, the loss component takes bank-specific historical loss data into account. In particular, similar BI and ATAL do not necessarily imply the same amount of regulatory capital, as ATAL_{10} and ATAL_{100} can be different, e.g., due to a heavier tailed loss distribution.

3.5 AMA Modeling Issues

At this point, we want to dwell on the AMA problem. One is inclined to ask why modeling OpRisk has turned out to be a serious issue that the Basel Committee is now planning to shelve soon. A variety of different aspects can be put forth to answer this question.

With the implementation of Basel II and its introduction of measuring operational risk, many financial institutions faced the problem of data scarcity. Naturally, the concept of OpRisk had been known before that point in time, but due to the non-consideration in terms of legal requirements, banks often did not possess a sufficiently large operational loss database to perform reasonable statistical analyses with. Particularly, as one is interested in high quantiles, requiring an even larger number of observations to obtain the same level of precision. The Basel Committee intended to solve this problem by demanding the banks to include external data into their models until the internal database is large enough to guarantee an acceptable OpRisk modeling process without additional data material from

other banks.

The inclusion of external data, however, entailed new issues for the practicioners. It was neither clear where to take the external data from, nor did the BCBS specify how these data need to be adjusted in order to complement the internal data accurately as far as the bank's individual risk profile is concerned. Bolancé et al. [2012, pp. 14-15] discuss three types of sources for external operational loss data and their characteristics. They also respond to the question of how to customise the external data such that they are in accordance with the internal data.

> "[...], we note that the size of the losses is very different in the internal and the external samples. The y-axis for internal losses ranges from 0.0 to 1.1, whereas the y-axis for the external data plot ranges from 0.0 to about 35.0. The reason for the different scales is that data from the external sample have been obtained from companies that may be larger than the company form which internal data were obtained. The visual effect is that both data sets have a similar coefficient of variation, so this means that a rescaling procedure would solve the problem of combining both data sets.", Bolancé et al. [2012, p. 15].

Another crucial point is the level at which the risk is measured. The 99.9% Value-at-Risk can be interpreted as the amount of operational losses that is on average exceeded only once in 1000 years. Not only is there no historical data for such a long time period, but even if there was, the operational risks of centuries ago (e.g. Black Death) are hardly comparable with the ones banks are facing today (e.g. cyber attacks). On that account, one needs to fit models to today's data and extrapolate in order to get sound estimates for the tail quantiles. In the AMA framework, the latter are usually obtained as the sample quantiles of simulated data from the 12-months compounded loss process (i.e., the distribution resulting from combining loss frequency and loss severity). In order to illustrate why estimating tail quantiles becomes a challenging task, we need the following result from probability theory that we state without a formal proof.

Theorem 1. *(Ruppert [2011, p. 50]) Let $Y_1, \ldots, Y_n$ be an i.i.d. sample with a cdf F. Suppose that F has a density f that is continuous and positive at $F^{-1}(q), 0 < q < 1$. Then for large n, the q-th sample quantile is approximately normally distributed with mean equal to the population quantile $F^{-1}(q)$ and variance equal to*

$$\frac{q(1-q)}{n\left[f\{F^{-1}(q)\}\right]^2}.$$

The interesting part in this theorem is the quadratic density in the denominator. Ignoring the numerator term for a moment, the sample quantile estimate exhibits the highest asymptotic variance in low density domains. As an example, Figure 3.2 shows the asymptotic sample quantile variance for a lognormal distribution. Noting the log-scale of the y-axis, one gets an idea of how imprecise the quantile estimation becomes in the right tail. Now how is this related to the 99.9% VaR for OpRisk? Firstly, the Value-at-Risk at level p coincides with the p-quantile of the loss distribution. Secondly, a lognormal distribution is a common choice for modeling OpRisk. So the conclusions above can be translated one-to-one into the operational risk setting.

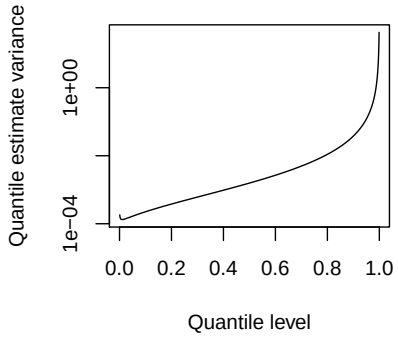

Figure 3.2: Asymptotic sample quantile variance for a lognormal distribution

Another point of critique is related to the risk measure Value-at-Risk itself. Due to the lack of sub-additivity, the VaR is not a coherent risk measure. For this reason, adding the VaRs for all business line/event type units does not guarantee a robust upper bound for the risk of the total risk. This part will be addressed in more detail in Sections 4.5 and 5.5.

One major issue with operational risk is the problem of insufficient data quality. Operational losses are often connected with the failure of individual departments or even single bank employees, as the Société Générale example shows. For fear of losing their jobs or suffering reputational damage, one may act on the assumption that a certain fraction of losses is not reported. This type of underreporting entails two major consequences from a modeling point of view. On the one hand, the earlier mentioned problem of data scarcity is intensified by this behaviour. On the other hand, the loss frequency (and possibly even the loss severity) is biased, clearly underestimating the true risk of operational losses. This problem can be remedied by introducing an underreporting function as discussed in

Bolancé et al. [2012].

Summary

The introduction of a capital charge for OpRisk in the second Basel Accord has increased the awareness of this type of risk considerably. Unlike other risks, Operational losses feature a number of undesirable characteristics which make the process of modeling, controlling and monitoring a challenging issue. The approaches presented in Basel II fail to provide a simple, risk-sensitive solution. To top it all, the high expectations regarding the AMA have not been satisfied yet. The new Standardised Measurement Approach shall combine risk-sensitivity and comparability between banks.

Chapter 4

Theory

The modeling of OpRisk requires a multitude of sophisticated statistical tools. It is the purpose of this chapter to give a brief outline of some of the concepts, which will be used in the applied part of this thesis.

4.1 Distributions

The decisive question in the context of risk management is which distribution the losses within a pre-specified time horizon do follow. In the application of OpRisk this question is typically subdivided into two parts, the frequency distribution and the loss severity distribution. Thanks to certain characteristics of operational loss data (e.g., positivity and heavy tails), the number of suitable parametric distributions can be reduced to a comparatively small number.

Distributions for Frequency Modeling

First, appropriate distributions for modeling the loss frequency within a 1 year horizon are considered. The number of operational loss events is clearly a non-negative integer with no natural upper bound. For this reason, a natural and popular choice is the Poisson distribution.

Definition 2 (Poisson Distribution)**.** *A random variable X is said to be Poisson distributed with intensity $\lambda > 0$, denoted $X \sim Poi(\lambda)$, if its probability mass function is given by*

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad k = 0, 1, 2, \ldots$$

While the Poisson distribution features the property of being a very simple model, one can argue it is simplistic. One of the key restrictions of the Poisson distribution is the equality of expected value and variance. The Negative Binomial distribution allows a relaxation of this implicit assumption and adds more modeling flexibility.

Definition 3 (Negative Binomial Distribution)**.** *A random variable X follows a Negative Binomial distribution with parameters $r > 0$ and $p \in (0, 1)$, $X \sim NB(r, p)$, if its probability mass function is*

$$\mathbb{P}(X = k) = \binom{k + r - 1}{k} p^k (1 - p)^r, \qquad k = 0, 1, 2, \ldots$$

From reading numerous publications on this subject, we got the impression that the Poisson distribution is the preferable model, since the benefits of the Negative Binomial distribution often do not outweigh the additional model complexity. For the underlying data in this thesis, however, neither of these distributions can be considered as the data material consists of aggregated monthly losses rather than individual losses.

Distributions for Severity Modeling

As far as the loss severity distribution is concerned, other characteristics need to be considered in the selection of reasonable parametric families. Unlike the loss frequency, the loss severity can take non-integer values allowing the use of continuous distribution functions. It is a common feature of OpRisk data to observe a large number of small losses and a small number of large losses. Yet, a simple exponential distribution is in general too light tailed to model the large losses accurately. We present five distributions that take the heavy-tailedness into consideration.

Definition 4 (Gamma Distribution)**.** *A random variable X is said to have a Gamma distribution with parameters $\alpha > 0$ and $\beta > 0$, denoted by $X \sim \Gamma(\alpha, \beta)$, if its probability density function is given by*

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha - 1} e^{-\beta x}.$$

The parametrisation is called the shape-rate parametrisation with shape α and rate β.

Definition 5 (Weibull Distribution)**.** *A random variable X is said to follow a Weibull distribution with parameters $\lambda \in (0, \infty)$ and $k \in (0, \infty)$, denoted by $X \sim WB(\lambda, k)$, if its*

cumulative distribution function is given by

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

λ and k are referred to as the scale and shape parameter, respectively.

Definition 6 (Lognormal Distribution)**.** *A random variable X is said to be lognormal distributed with location $\mu \in \mathbb{R}$ and scale parameter $\sigma > 0$, denoted by $X \sim \mathcal{LN}(\mu, \sigma^2)$, if its probability density function is given by*

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-[\log(x)-\mu]^2/(2\sigma^2)}.$$

The Gamma, Weibull and Lognormal distribution are the most popular choices when it comes to modeling loss severity. In the Peaks over Threshold approach in Extreme Value Theory these regular distributions are complemented by a Generalised Pareto distribution that is fitted to the right tail.

Definition 7 (Generalised Pareto Distribution)**.** *The cumulative distribution function of the GPD is given by*

$$G_{\xi,\beta}(x) = \begin{cases} 1 - \left(1 + \xi\frac{x}{\beta}\right)^{-1/\xi}, & \xi \neq 0 \\ 1 - \exp\left(-\frac{x}{\beta}\right), & \xi = 0, \end{cases} \tag{4.1}$$

where $\beta > 0$, and $x \geq 0$ when $\xi \geq 0$ and $0 \leq x \leq -\beta/\xi$ when $\xi < 0$. ξ and β are called the shape and scale parameter, respectively. (See McNeil et al. [2005, p. 275])

The mean of a random variable X having Generalised Pareto distribution is

$$\mathbb{E}[X] = \frac{\beta}{1-\xi}. \tag{4.2}$$

The details of this result are attached in the Appendix. As modeling with two different distributions adds intricacy, another distribution is considered. The Generalised Champernowne distribution was recently put forward in Bolancé et al. [2012] to be a universal distribution class that combines the behaviour of a lognormal distribution in the body and the behaviour of a GPD in the tail.

Definition 8 (Generalised Champernowne Distribution)**.** *The cumulative distribution func-*

tion of a generalised Champernowne distribution is given by

$$F_{\alpha,M,c}(x) = \frac{(x+c)^\alpha - c^\alpha}{(x+c)^\alpha + (M+c)^\alpha - 2c^\alpha}, \qquad x \geq 0,$$

where $\alpha > 0, M > 0$ *and* $c \geq 0$.

An explicit expression for the quantile function $F^{-1}_{\alpha,M,c}$ can be found regularly by solving

$$p = F_{\alpha,M,c}(F^{-1}_{\alpha,M,c}(p))$$

for $F^{-1}_{\alpha,M,c}(p)$. One obtains

$$F^{-1}_{\alpha,M,c}(p) = \left(\frac{p[(M+c)^\alpha - 2c^\alpha] + c^\alpha}{1-p}\right)^{1/\alpha} - c, \qquad p \in (0,1).$$

This step will be necessary when it comes to Quantile-Quantile plots in Chapter 5, as this distribution is not considered as one of the standard distributions in statistical modeling and, hence, is not implemented as a default in R.

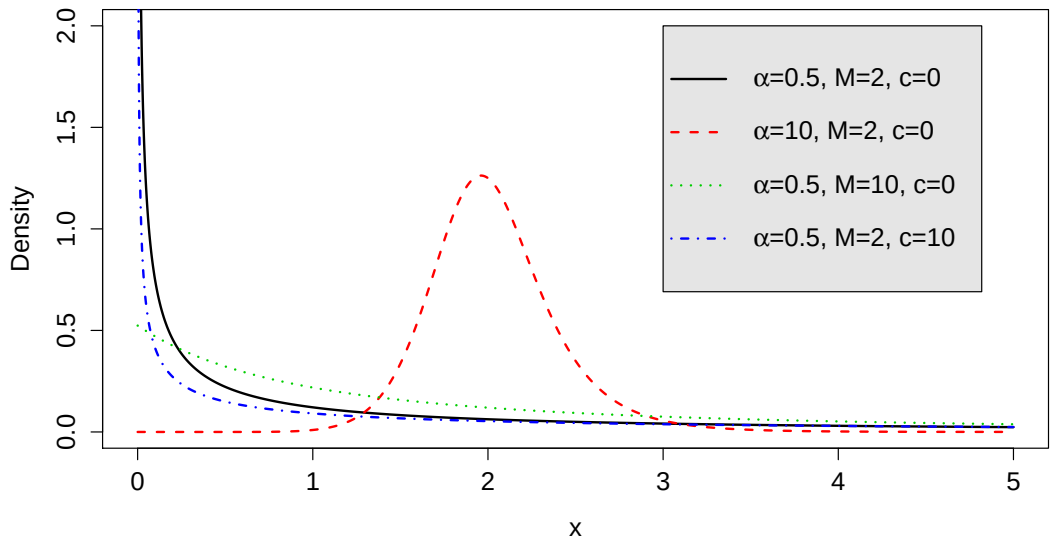

Figure 4.1: Density functions of a Generalised Champernowne distribution with different parameters

4.2 Copulas

Once univariate models have been fitted to different parts of the data (e.g., different event types), one should ask if these data are correlated in some sense and hence, need to be modeled jointly. This section is devoted to this question introducing the concept of copulas. Copulas enable practicioners to deal with dependence in a convenient way, as they allow them to model the marginal distributions and the dependence individually. This short introduction is based on the more extensive descriptions of McNeil et al. [2005, pp. 185-237] and Nelsen [2007].

Definition 9 (Copula)**.** *A distribution function on $[0,1]^d$ with standard uniform marginal distributions is called a d-dimensional copula.*

Theorem 2. *For all copulas $C : [0,1]^d \to [0,1]$ it holds*

1. *$C(u_1, \ldots, u_d)$ is increasing in each component u_i.*
2. *$C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$ for all $i \in \{1, \ldots, d\}$, $u_i \in [0,1]$.*
3. *For all $(a_1, \ldots, a_d), (b_1, \ldots, b_d) \in [0,1]^d$ with $a_i \leq b_i$ we have the rectangle inequality*

$$\sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1+\cdots+i_d} C(u_{1i_1}, \ldots, u_{di_d}) \geq 0,$$

where $u_{j1} = a_j$ and $u_{j2} = b_j$ for all $j \in \{1, \ldots, d\}$.

The main result in the context of copulas is called Sklar's Theorem and describes the one to one relationship between multivariate distribution functions and copulas applied to the marginal distributions. It is of utmost importance when it comes to the justification why copulas have become so popular in financial modeling.

Theorem 3 (Sklar)**.** *Let F be a joint distribution function with margins $F_1, \ldots, F_d$. Then there exists a copula $C : [0,1]^d \to [0,1]$ such that, for all $x_1, \ldots, x_d \in [-\infty, \infty]$,*

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)). \tag{4.3}$$

If the margins $F_1, \ldots, F_d$ are continuous, then C is unique; otherwise C is uniquely determined on $RanF_1 \times RanF_2 \times \cdots \times RanF_d$, where $RanF_i = F_i([-\infty, \infty])$ is the range of F_i. Conversely, if C is a copula and $F_1, \ldots, F_d$ are univariate distribution functions, then the function F defined in (4.3) is a joint distribution function with margins $F_1, \ldots, F_d$.

Proof. See McNeil et al. [2005, p. 187]. □

Given the joint distribution function F with margins $F_1, \ldots, F_d$, Sklar's Theorem says that the corresponding copula is found to be

$$C(u_1, \ldots, u_d) = F(F_1^{\leftarrow}(u_1), \ldots, F_d^{\leftarrow}(u_d)), \tag{4.4}$$

where $F_i^{\leftarrow}(x) \triangleq \inf\{y : F_i(y) \geq x\}$ is the generalised inverse function.

Elliptical Copulas

Having reviewed some of the basic properties of copulas, we continue by introducing the most popular copulas. We begin by highlighting the family of elliptical copulas containing both, the Gaussian copula and Student's t-copula. They can easily be derived from (4.4) by assuming Gaussian and t-margins, respectively.

Definition 10 (Gauß Copula)**.** *The Gaussian copula is defined by*

$$C^{Ga}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(u_1, \ldots, u_d) = \int_{-\infty}^{\phi_1^{-1}(u_1)} \cdots \int_{-\infty}^{\phi_d^{-1}(u_d)} \frac{1}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}|}} \exp\left\{(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right\} dx_d \cdots dx_1,$$

where $\boldsymbol{x} = (x_1, \ldots, x_d)$, $\boldsymbol{\mu} \in \mathbb{R}^d$, $\boldsymbol{\Sigma} \in \mathbb{R}^{d \times d}$ *with* $|\boldsymbol{\Sigma}| > 0$ *and* ϕ_d^{-1} *is the quantile function of a normal distribution with mean* μ_i *and variance* σ_{ii}*.*

Definition 11 (Student's t-Copula)**.** *Student's t-copula is defined by*

$$C^{t}_{\nu,\boldsymbol{\Sigma}}(u_1, \ldots, u_d) = \int_{-\infty}^{t_1^{-1}(u_1)} \cdots \int_{-\infty}^{t_d^{-1}(u_d)} \frac{\Gamma[(\nu+d)/2](\nu\pi)^{-d/2}}{\Gamma(\nu/2)|\boldsymbol{\Sigma}|^{1/2}\left[1 + \frac{1}{\nu}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]^{(\nu+d)/2}} dx_d \cdots dx_1,$$

where $\boldsymbol{x} = (x_1, \ldots, x_d)$, $\boldsymbol{\mu} \in \mathbb{R}^d$, $\boldsymbol{\Sigma} \in \mathbb{R}^{d \times d}$ *with* $|\boldsymbol{\Sigma}| > 0$ *and* t_i^{-1} *the quantile function of a t distribution with mean* μ_i*, variance* σ_{ii} *and* ν_i *degrees of freedom.*

The name *elliptical copulas* refers to the shape of these types of copulas. Due to the quadratic form $(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu})$ the contour plot (Figure 4.2) consists solely of ellipsoids.

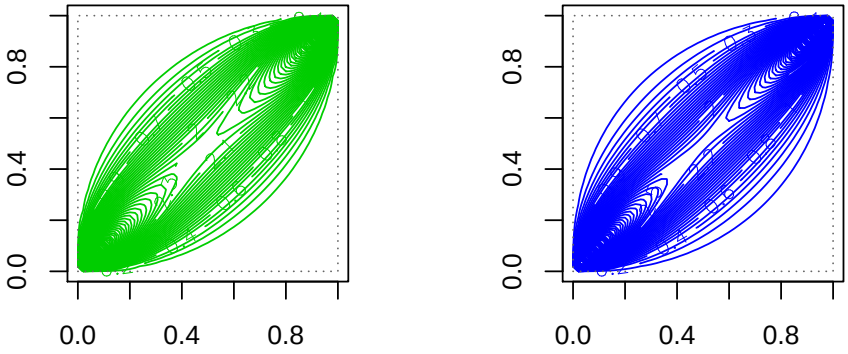

Figure 4.2: Probability density contour plots of the bivariate Gaussian copula ($\rho = 0.9$) and the bivariate t-copula ($\rho = 0.9, \nu = 8$).

Archimedean Copulas

Another large group of copulas is given by the family of Archimedean copulas. In contrast to the Gauß and t copula, Archimedean copulas are generated by a so-called generator function $\phi : [0,1] \to [0,\infty]$ that is assumed to be decreasing and to satisfy the condition $\phi(1) = 0$. If also $\phi(0) = \infty$, ϕ is said to be a *strict* generator, which will be a necessary condition to define Archimedean copulas in dimensions $d \geq 3$. ϕ generates a copula by

$$C(u_1, \ldots, u_d) = \phi^{-1}(\phi(u_1) + \cdots + \phi(u_d)),$$

which usually yields a nicer expression than the ones obtained for the elliptical family. Table 4.1 lists some of the most well-known Archimedean copulas. Their probability density contour plots are in general non-elliptical as Figure 4.3 shows.

Table 4.1: Most important bivariate Archimedean copulas

Name	**Generator**	θ **Range**	**Strict**	$C(u,v)$
Clayton	$\frac{1}{\theta}(t^{-\theta} - 1)$	$\theta \geq -1$	$\theta \geq 0$	$\left[\max\left\{u^{-\theta} + v^{-\theta} - 1, 0\right\}\right]^{-1/\theta}$
Frank	$-\log\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right)$	$\theta \in \mathbb{R}$	Yes	$-\frac{1}{\theta}\log\left[1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right]$
Gumbel	$(-\log t)^{\theta}$	$\theta \geq 1$	Yes	$\exp\left[-\left\{(-\log(u))^{\theta} + (-\log(v))^{\theta}\right\}^{1/\theta}\right]$

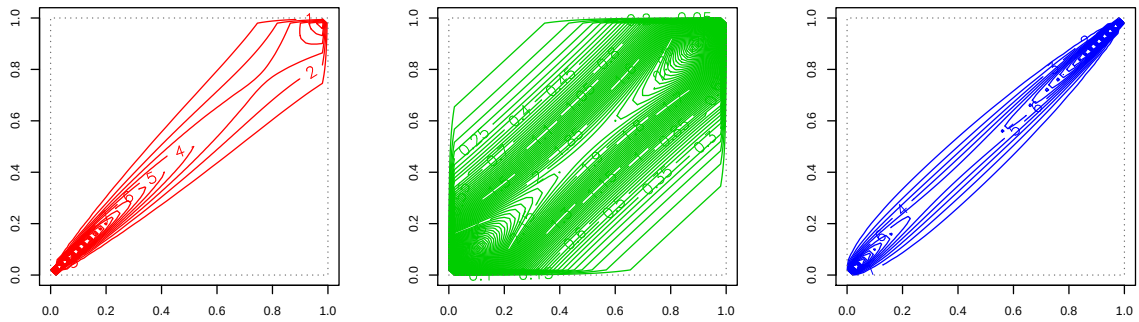

Figure 4.3: Probability density contour plots of the bivariate Clayton copula (left), Frank copula (center) and Gumbel copula (right) with $\theta = 8$.

Measures of Dependence

The concept of copulas is motivated by the idea of modeling dependences between different data samples. Therefore, it is closely related to the question of how dependence can be measured. In this section, Kendall's τ and Spearman's ρ_S are introduced. We stress their benefits compared to Pearson's linear correlation coefficient ρ, which is by far the most popular measure of dependence. On a related note, we also present the concept of tail dependence.

When we speak of the correlation between two random variables X and Y, it is mostly the linear correlation coefficient

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}$$

that is spoken of. Besides linear correlation, there is a less popular concept called rank correlation. To illustrate the differences between these types of correlation, we consider Figure 4.4. Ten random numbers were drawn from two i.i.d. exponential random variables X and Y. The resulting scatterplot (left plot) indicates there is hardly any correlation between X and Y. Yet, if the observations are logarithmised, the empirical linear correlation, i.e., the slope of the least squares line, changes drastically from $\hat{\rho}(X,Y) = 0.00689$ to $\hat{\rho}(\log(X), \log(Y)) = 0.8955$ and suggests a strong positive dependence between $\log(X)$ and $\log(Y)$. On the other hand, if we compare the ranks of X and $\log(X)$ as well as the ranks of Y and $\log(Y)$, we find that they haven't changed since the logarithm is a positive monotone transformation, i.e. the first derivative is strictly positive. For this reason, every rank correlation measure remains unchanged under this transformation.

The latter observation is closely linked to the concept of concordance. Given two

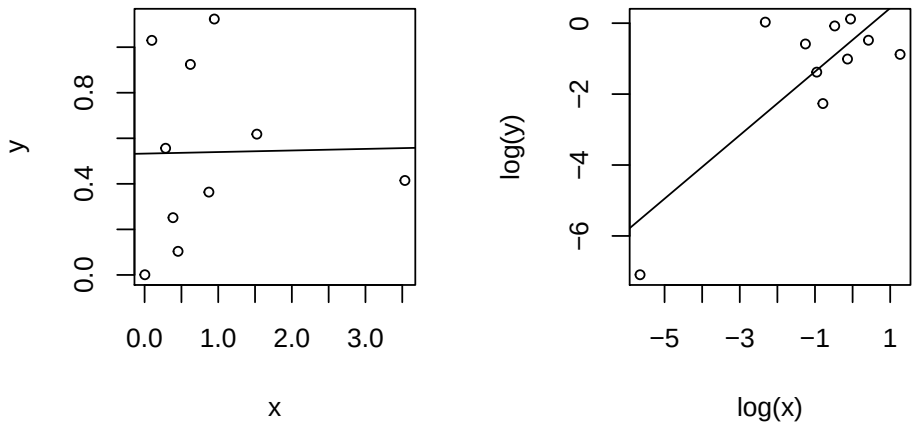

Figure 4.4: Scatterplots of two independent Exp(1) random variables in original scale (left) and after log-transformation (right). The corresponding least squares line shows the linear correlation.

points in $\mathbb{R}^2$, (x_1, x_2) and $(\tilde{x}_1, \tilde{x}_2)$, they are called concordant if $(x_1 - \tilde{x}_1)(x_2 - \tilde{x}_2) > 0$ and discordant if $(x_1 - \tilde{x}_1)(x_2 - \tilde{x}_2) < 0$. Putting this back into a probabilistic context, we define the concordance function as

$$Q = \mathbb{P}\left((X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) > 0\right) - \mathbb{P}\left((X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) < 0\right),$$

where $(X_1, X_2) \sim F$ and $(\tilde{X}_1, \tilde{X}_2) \sim \tilde{F}$. Letting $F = \tilde{F}$, we obtain Kendall's τ.

Definition 12 (Kendall's τ)**.** *Let (X_1, X_2) and $(\tilde{X}_1, \tilde{X}_2)$ be two independent random vectors with the same distribution, then Kendall's τ is defined as the concordance function*

$$\tau(X_1, X_2) = \mathbb{P}\left((X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) > 0\right) - \mathbb{P}\left((X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) < 0\right).$$

An unbiased estimator is given by

$$\tilde{\tau}(X_1, X_2) = \frac{c - d}{\binom{n}{2}},$$

where c and d are the number of concordant and discordant observations, respectively, and n is the total number of bivariate observations.

Spearman's rho can be defined in multiple different ways. The simplest definition involves Pearson's linear correlation coefficient ρ.

Definition 13 (Spearman's ρ_S)**.** *Spearman's rho is given by*

$$\rho_S(X_1, X_2) = \rho(F_1(X_1), F_2(X_2)),$$

where X_1 and X_2 are random variables with marginal distribution functions F_1 and F_2. An unbiased estimator is given by

$$\tilde{\rho}_S(X_1, X_2) = \rho(\hat{F}_1(X_1), \hat{F}_2(X_2)),$$

where $\hat{F}_1$ and $\hat{F}_2$ are the empirical cdfs of X_1 and X_2, defined by $\hat{F}(x_i) = \frac{Rank(x_i)}{n+1}$.

It can be shown that both, Kendall's τ and Spearman's ρ_S, depend solely on the corresponding copula. Since copulas are invariant under increasing transformations, these two dependence measures do not change in the example of Figure 4.4.

All three measures ρ, τ and ρ_S have in common that they focus on the overall dependence between X_1 and X_2 rather than putting more weight on some specific observations. However, in risk management one is particularly interested in upper tail events. If huge losses of X_1 and X_2 happen to coincide regularly, this needs to be considered even if the remaining losses are hardly correlated with each other, suggesting an overall independence. This thought motivates the use of another dependence measure, the tail dependence.

Definition 14 (Upper tail dependence)**.** *Let $(X, Y) \sim F$ be a random vector with marginal cdfs F_1 and F_2. The upper tail dependence is defined as*

$$\lambda_U = \lim_{u \to 1} \mathbb{P}(Y > F_2^{\leftarrow}(u) | X > F_1^{\leftarrow}(u)).$$

For Archimedean copulas it can be shown that

$$\lambda_U = 2 - \lim_{x \searrow 0} \frac{1 - \phi^{[-1]}(2x)}{1 - \phi^{[-1]}(x)},$$

where $\phi^{[-1]}$ refers to the pseudo-inverse of the generator ϕ, defined by

$$\phi^{[-1]}(x) = \begin{cases} \phi^{-1}(x) & 0 \leq x \leq \phi(0) \\ 0 & \phi(0) \leq x \leq \infty. \end{cases}$$

Following this result from Nelsen [2007, Corollary 5.4.3], some upper tail dependences are listed in Table 4.2. For the tail dependence of a t-copula, Demarta and McNeil [2005,

Proposition 1] was used.

Table 4.2: Upper tail dependence

	Clayton	**Frank**	**Gumbel**	**Gauß**	**Student's t**
λ_U	0	$2 - 2^{1/\theta}$	0	0	$t_{\nu+1}(-\sqrt{\nu+1}\sqrt{1-\rho}/\sqrt{1+\rho})$

Fitting a Copula

The procedure of estimating the parameters within the copula approach can be divided into three methods, Full Maximum Likelihood (FML), a two-stage estimation method based on an Inference Functions for Margins (IFM) and finally, a non-parametric approach called Canonical Maximum Likelihood (CML). A short outline of these methods is given based on Yan et al. [2007], in which a slightly more detailed presentation in the context of the R-package *copula* can be found.

The FML estimation method is a parametric approach that involves a simultaneous estimation of copula-based parameters α and marginal parameters β. Let F_i and f_i denote the cdf and pdf of the marginal distributions, respectively, where $i = 1, \ldots, d$. If c refers to the copula density and n d-dimensional realisations $\boldsymbol{x_i} = (x_{i1}, \ldots, x_{id})$ are given, then the loglikelihood

$$\log \mathcal{L}(\alpha, \beta) = \prod_{i=1}^{n} \log c \left\{ F_1(x_{i1}; \beta), \ldots, F_d(x_{id}; \beta); \alpha \right\} + \sum_{i=1}^{n} \sum_{j=1}^{d} \log f_i(x_{ij}; \beta)$$

is maximised with respect to both, α and β.

Due to numerical instability of the FML method, another approach was developed in Joe and Xu [1996] that is called IFM. Contrary to the FML, the IFM prioritises the estimation of the vector β of marginal parameters

$$\hat{\beta}_{j,IFM} = \arg\max \sum_{i=1}^{n} \log f_i(x_{ij}; \beta_j), \qquad j = 1, \ldots, d.$$

In a second step, the vector of copula-based parameters α is estimated by

$$\hat{\alpha}_{IFM} = \arg\max \sum_{i=1}^{n} \log c \left\{ F_1(x_{i1}; \hat{\beta}_{IFM}), \ldots, F_d(x_{id}; \hat{\beta}_{IFM}); \alpha \right\}.$$

Note, however, that both, FML and IFM, are parametric estimation methods and may lead to poor results if the marginal distributions are not specified correctly. If one is mainly interested in obtaining a consistent estimate of the dependence, the non-parametric CML method comes into play, which solves the optimisation problem

$$\hat{\alpha}_{CML} = \arg\max \sum_{i=1}^{n} \log c \left\{ u_{i1}, \ldots, u_{id}; \alpha \right\},$$

where $(u_{i1}, \ldots, u_{id})$ is the vector obtained from transforming $(x_{i1}, \ldots, x_{id})$ into uniform pseudo observations.

Goodness-of-Fit for Copulas

Knowing the most important copula families and having fitted different models to the data, one faces two different questions. Which fitted copula model is the best one on a relative scale? And how does this model perform in absolute terms, i.e., is it possible to test whether the best relative model also fits the data? These questions turn out to be far from trivial. A detailed comparison of goodness-of-fit methods for copulas can be found in Genest et al. [2009]. In the following, we outline parts of the content of this paper.

Given independent copies $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ of a continuous random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ with joint cdf H and marginal distributions $F_1, \ldots, F_d$, Sklar's Theorem yields the copula representation $H(x_1, \ldots, x_d) = C\left(F_1(x_1), \ldots, F_d(x_d)\right)$ with a unique C. It shall be tested whether C belongs to a specific parametric class

$$\mathcal{C}_0 = \{C_\theta : \theta \in \mathcal{O}\}$$

with $\mathcal{O} \in \mathbb{R}^p$ being an open set for some integer $p \geq 1$. The empirical copula is defined as

$$C_n(\boldsymbol{u}) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(U_{i1} \leq u_1, \ldots, U_{id} \leq u_d),$$

where $\boldsymbol{u} = (u_1, \ldots, u_d) \in [0,1]^d$ and $\boldsymbol{U}_1 = (U_{11}, \ldots, U_{1d}), \ldots, \boldsymbol{U}_n = (U_{n1}, \ldots, U_{nd})$ are rank-based pseudo-observations of $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, i.e., $U_{ij} = R_{ij}/(n+1)$. In this definition, R_{ij} is the rank of X_{ij} within the vector $\boldsymbol{X}_j = (X_{1j}, \ldots, X_{nj})$. Denoting an estimate of θ based on these pseudo-observations by $\theta_n = \mathcal{T}_n(\boldsymbol{U}_1, \ldots, \boldsymbol{U}_n)$, the process

$$\mathbb{C}_n = \sqrt{n}(C_n - C_{\theta_n})$$

is defined, being a measure of distance between the empirical copula and the parametric copula of H_0. The test statistics we consider in this thesis is

$$S_n = \int_{[0,1]^d} \mathbb{C}_n(\boldsymbol{u})^2 dC_n(\boldsymbol{u}).$$

Two other test statistics mentioned by Genest et al. [2009] are based on Rosenblatt's transform $\mathcal{R}$, exploiting the equivalence of $H_0 : \mathcal{U} \triangleq (F_1(X_1), \ldots, F_d(X_d)) \sim C$ with $H_0^* : \mathcal{R}_{\theta_n}(\mathcal{U}) \sim C_\perp$, where $C_\perp$ denotes the independent copula. These test statistics are defined as

$$S_n^{(C)} = n \int_{[0,1]^d} \left(D_n(\boldsymbol{u}) - C_\perp(\boldsymbol{u})\right)^2 dD_n(\boldsymbol{u})$$

and

$$S_n^{(B)} = n \int_{[0,1]^d} \left(D_n(\boldsymbol{u}) - C_\perp(\boldsymbol{u})\right)^2 d\boldsymbol{u}.$$

In the extensive study of different Goodness-of-Fit tests in Genest et al. [2009], these three tests have proven to be reliable tests. Furthermore, Appendix A and D of the same paper contain the algorithms for approximate p-values of these tests. They are implemented in the *R copula* package.

4.3 Extreme Value Theory

This section aims to give a brief introduction to classical Extreme Value Theory highlighting the main result, which is the Fisher-Tippet-Gnedenko theorem. To obtain a more thorough angle on this subject, we refer to Embrechts et al. [1997] on whose detailed analysis this outline is based.

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. non-degenerate random variables with common cdf F. Classical EVT strives for making statements on how the most extreme values of this sequence behave. More precisely, one is interested in assessing the distribution of the minimum $m_n = \min(X_1, \ldots, X_n)$ and/or the maximum $M_n = \max(X_1, \ldots, X_n)$, $n \in \mathbb{Z}_+$. In the sequel, we will without loss of generality focus on the maximum case.

A natural first approach is to study the distribution function. Basic probabilitistic methods yield

$$\mathbb{P}(M_n \leq x) = \mathbb{P}(X_1 \leq x, \ldots, X_n \leq x) \overset{iid}{=} \prod_{i=1}^{n} \mathbb{P}(X_i \leq x) = F^n(x). \tag{4.5}$$

This result per se is not helpful to model extreme events since it can be shown that the limiting distribution G is a degenerate function in $x_F = \sup\{x \in \mathbb{R} : F(x) < 1\}$, i.e.

$$\lim_{n\to\infty} F^n(x) = G(x) = \begin{cases} 0, & x < x_F \\ 1, & x \geq x_F. \end{cases} \tag{4.6}$$

In line with this result, one concludes that $M_n \xrightarrow{P} x_F$ and due to the maximum being non-decreasing, it even holds $M_n \xrightarrow{a.s.} x_F$. This information might be of use when estimating x_F by the maximum of a number of i.i.d. data observations, but for stochastic purposes it is little help.

To avoid getting a non-degenerate limiting distribution, it is – similarly to the Central Limit Theorem – necessary, to standardise the quantity of interest. The Extremal Types Theorem (also known as Fisher-Tippett-Gnedenko Theorem, see Fisher and Tippett [1928] and Gnedenko [1943]) then yields the first important convergence result stating that the distribution of the normalised maximum of an i.i.d. sequence belongs to one out of three families. We use the version mentioned in Coles et al. [2001, p. 46].

Theorem 4 (Extremal Types)**.** *Let $X_1, \ldots, X_n$ be a sequence of independent and identically-distributed random variables and let $M_n \triangleq \max\{X_1, \ldots, X_n\}$. If there exist sequences of constants $\{\mu_n\}$ and $\{\beta_n > 0\}$ such that*

$$\mathbb{P}\left(\frac{M_n - \mu_n}{\beta_n} \leq x\right) \to G(x) \quad \text{as } n \to \infty,$$

where G is a non-degenerate distribution function, then G belongs to one of the following families:

$$\textit{I (Gumbel):}\ G(x) = \exp\left\{-\exp\left[-\left(\frac{x-\mu}{\beta}\right)\right]\right\}, \quad -\infty < z < \infty;$$

$$\textit{II (Fréchet):}\ G(x) = \begin{cases} 0, & \textit{if } x \leq \mu, \\ \exp\left\{-\left(\frac{x-\mu}{\beta}\right)^{-\xi}\right\}, & \textit{otherwise;} \end{cases}$$

$$\textit{III (Reversed Weibull):}\ G(x) = \begin{cases} \exp\left\{-\left[-\left(\frac{x-\mu}{\beta}\right)^{\xi}\right]\right\}, & \textit{if } x < \mu \\ 1 & \textit{otherwise.} \end{cases}$$

for parameters $\beta > 0$, μ and, in the case of families II and III, $\xi > 0$.

These three families of extreme value distributions are summarised to the so-called Generalised Extreme Value distribution (GEV)

$$G_{\mu,\beta,\xi}(x) = \begin{cases} \exp\left[-\exp(-\frac{x-\mu}{\beta})\right], & \xi = 0, \\ \exp\left[-(1+\xi\frac{x-\mu}{\beta})^{-1/\xi}\right], & \xi \neq 0, \end{cases} \tag{4.7}$$

where $1 + \xi\frac{x-\mu}{\beta} > 0$. This is often referred to as the Jenkinson-von Mises representation (see Von Mises [1936] and Jenkinson [1955]).

Peaks Over Threshold

The most popular approach to deal with EVT is called Peaks Over Threshold (POT). This method uses all observations exceeding some threshold level u, which is in contrast to the so-called block maxima approach that considers the maxima within prespecified time intervals. The POT approach features the advantage of not wasting extreme observations caused by the fact that an even larger value was observed within the same time interval. But just like in the block maxima case, where fixing the size of the time intervals is one major issue, determining the threshold level turns out to be the Achilles heel of POT.

The idea of POT is based on the following theoretical result. Let again $X_1, \ldots, X_n$ be an i.i.d. sequence of degenerate rvs with common cdf F and let X denote an arbitrary term in this sequence. If the requirements for the Extremal Types theorem are fulfilled, then $\mathbb{P}(M_n \leq x) \approx G(x)$ for n large, where G is GEV (4.7). Fixing some threshold u, the conditional distribution of the exceedances $X - u$ given that $X > u$ can be calculated as

$$\mathbb{P}(X - u > x | X > u) = \frac{1 - F(u+x)}{1 - F(u)} \approx \left[1 + \xi\frac{x}{\tilde{\beta}}\right]^{-1/\xi}, \tag{4.8}$$

where

$$\tilde{\beta} = \beta + \xi(u - \mu). \tag{4.9}$$

Details of this calculation can be found in Coles et al. [2001, pp. 75-77] and rely on the first order Taylor approximation $\log F(z) \approx -(1 - F(z))$, which holds true only if $F(z) \approx 1$.[1] In particular, this implicates that the approximation (4.8) is only valid if $F(u) \approx 1$, or equivalently if the threshold u is large enough with respect to the support of F. The

[1] Please note the typo in equation (4.7) in Coles et al. [2001, p. 77]. The expression on the right hand side in the second line should read $\left[1 + \frac{\xi(y/\sigma)}{1+\xi(u-\mu)/\sigma}\right]^{-1/\xi}$.

distribution in (4.8) is known as the Generalised Pareto Distribution (GPD).

4.3.1 Threshold Selection

As can be seen in the previous paragraph, the introduction of the threshold u is rather vague and one might ask in which cases the threshold is in fact large enough. Intuitively, it is argued a low threshold would increase the number of observations exceeding u resulting in a higher number of data points contributing to model estimation and reducing the parameter uncertainty. On the other hand, a low threshold might introduce bias because the approximation in (4.8) does not apply to this choice of threshold. If, in contrast, the threshold is chosen very high, we can almost be certain about the unbiasedness of our estimates. The price we pay is the higher parameter estimates variance. The problem of threshold selection can therefore be reduced to a bias-variance trade-off.

Various tools are used to facilitate this decision to practicioners. In the following, we want to present the most common of them.

Definition 15 (Mean Excess function)**.** *Let X be some arbitrary random variable with finite mean. The mean excess function of X is given by*

$$e(u) = \mathbb{E}[X - u | X > u].$$

The mean excess function returns the expected value of the exceedances subject to different threshold choices. The importance of this function in this context traces back to the following result.

Theorem 5. *Let $X - u \sim GPD_{\xi,\beta(u)}$, where $\beta(u)$ is given by (4.9). Then the mean excess function is*

$$e(u) = \frac{\beta(u)}{1 - \xi}. \tag{4.10}$$

Proof. See equation (4.2). □

The detail which makes this result attractive for practical purposes is the linearity of $e(u)$ in regions where the exceedances are approximately GPD. One can perform a graphical analysis on the so-called sample mean excess plot (see McNeil et al. [2005, p. 279])

$$e_n(u) = \frac{\sum_{i=1}^{n}(X_i - u)I_{\{X_i > u\}}}{\sum_{i=1}^{n} I_{\{X_i > u\}}}, \tag{4.11}$$

i.e., an empirical estimate of the mean excess function. Due to the linearity of the latter for GPD exceedances, the threshold is selected as the lowest value u exhibiting a linear relation for all $u' > u$. It should be stressed, however, that the sample mean excess plot seldom looks entirely linear for higher thresholds, but fluctuates greatly since the value is built over few observations only. For this reason it can be advisable to omit the largest observations in the graphical analysis.

Another option is to consider the median excess plot, which is the analogue for the median. Due to the median being more robust to outliers, this plot is sometimes preferred to the mean excess plot (see, e.g., Rootzén and Tajvidi [1997]).

Definition 16 (Median Excess function)**.** *Let X be some arbitrary random variable with cdf F. The median excess function of X is given by*

$$e_{med}(u) = \mathbb{M}[X - u | X > u] = \inf\left\{x \in \mathbb{R} : \frac{F(x+u) - F(u)}{1 - F(u)} \geq 1/2\right\}.$$

We find the median excess function to be linear for GPD exceedances, too.

Theorem 6. *Let $X - u \sim GPD_{\xi,\beta(u)}$, where $\beta(u)$ is given by (4.9). Then the median excess function is*

$$e_{med}(u) = (2^{\xi} - 1)\frac{\beta(u)}{\xi}. \tag{4.12}$$

Proof. Equating the right hand side of (4.8) with 1/2. □

4.3.2 Parameter Estimation

Given exceedances $x_1, \ldots, x_n$ over some fixed threshold u, the parameters ξ and σ can be found by means of maximum likelihood estimation. In view of (4.1) one maximises the loglikelihood function

$$\log \mathcal{L}(\xi, \beta) = -n \log \beta - (1 + 1/\xi) \sum_{i=1}^{n} \log(1 + \xi x_i/\beta),$$

provided that $\xi \neq 0$. This has to be solved numerically. In the Gumbel case $\xi = 0$, the loglikelihood becomes

$$\log \mathcal{L}(\beta) = -n \log \beta - \sum_{i=1}^{n} x_i/\beta,$$

resulting in the ML estimate $\hat{\beta}_{ML} = n^{-1} \sum_{i=1}^{n} x_i$.

4.4 Goodness-of-Fit

In order to evaluate the accuracy of a model, numerous methods can be applied. Three techniques are presented in this section, as they will be used in the loss distribution approach in Chapter 5. The Q-Q plot and the tail plot relate data samples with theoretical probability distributions, with the tail plot placing more weight on the upper tail. These graphical methods are complemented by the formal Anderson-Darling test.

4.4.1 Quantile-Quantile Plot

The Quantile-Quantile (Q-Q) plot is a well-known graphical method that allows a comparison between a sample of data and a statistical population. If one wants to check whether a sample $x_1, \ldots, x_n$ follows a specific probability distribution F, the Q-Q plot is obtained from the points

$$\left\{(F^{-1}(k/(n+1)), x_{(k)}) : k = 1, \ldots, n\right\},$$

where $F^{-1}(p)$ refers to the p quantile of F and $x_{(k)}$ is the order statistic of rank k. Asymptotically, these points follow the 45° line $y = x$, provided that the sample $x_1, \ldots, x_n$ in fact follows F.

4.4.2 Tail Plot

The tail plot is another graphical tool to assess the quality of a model fit. For a given data sample $x_1, \ldots, x_n$ the points

$$\left\{\left(x_{(k)}, 1 - \hat{F}\left(x_{(k)}\right)\right) : k = 1, \ldots, n\right\}$$

are plotted on a log scale, where $\hat{F}(x_{(k)}) = k/(n+1)$ is the empirical cdf at order statistic k. These points are then compared with the line defined by

$$\{(x, 1 - F(x)) : x \in \mathbb{R}\},$$

relating all real values with their theoretical survival probabilities. If the sample was generated by F, then the points lie asymptotically on the line. The log-scaling highlights especially differences in the tail, which gave the plot its name.

4.4.3 The Anderson-Darling Test

To formally test if a sample of data $x_1, \ldots, x_N$ follows a specific probability distribution F, we put forward the Anderson-Darling test from Anderson and Darling [1952], which can be seen as a special case of the Cramér-von Mises test with focus on the tails. Under the null hypothesis, $x_1, \ldots, x_N$ describe a realisation of independent draws from F. In its simplest form the test statistic is given by

$$A^2 = -N - S,$$

where N is the number of observations and S is defined as

$$S = \frac{1}{N} \sum_{i=1}^{N} (2i-1) \left[\log\left(F(x_{(i)})\right) + \log\left(1 - F(x_{(N+1-i)})\right) \right] \tag{4.13}$$

with $x_{(1)} \leq \cdots \leq x_{(N)}$ being the ordered data. It follows from (4.13) that A^2 is infinite iff there exists at least one x_i with $F(x_i) \in \{0, 1\}$. Critical values for this test depend on the distribution function being tested and are only known exactly for the most popular families of distributions (see Stephens [1974] or D'Agostino [1986]). As far as extreme value distributions are concerned, we refer to Stephens [1977] for a disquisition on the Gumbel double exponential distribution and Choulakian and Stephens [2001] for the GPD case. If the exact critical value is not known, it can be estimated by means of Monte Carlo simulation.

4.5 Risk Measures

A natural question that arises when it comes to risk quantification is how to measure risk. And in fact, even though this question has been discussed for several decades, it is still apparent in today's risk management discussions. Risk managers have been striving for expressing risk in a simple figure for a long time. This figure should indicate the degree of risk exposure and fulfill several properties that are widely known as the risk measure axioms. We stress these axioms using the notation from McNeil et al. [2005].

Definition 17 (Coherent risk measure)**.** *A function $\varrho : \mathcal{M} \to \mathbb{R}$ on a convex cone $\mathcal{M}$ is called a coherent risk measure if it satisfies*

1. *For all $L \in \mathcal{M}$ and every $l \in \mathbb{R}$ it holds $\varrho(L + l) = \varrho(L) + l$. (Translation invariance)*

2. *For all* $L_1, L_2 \in \mathcal{M}$ *we have* $\varrho(L_1 + L_2) \leq \varrho(L_1) + \varrho(L_2)$. *(Subadditivity)*

3. *For all* $L \in \mathcal{M}$ *and every* $\lambda > 0$ *it holds* $\varrho(\lambda L) = \lambda \varrho(L)$. *(Positive homogeneity)*

4. *For* $L_1, L_2 \in \mathcal{M}$ *with* $L_1 \leq L_2$ *almost surely we have* $\varrho(L_1) \leq \varrho(L_2)$. *(Monotonicity)*

In the following we introduce the two most common measures of risk, the Value-at-Risk and the Expected Shortfall and discuss their benefits as well as their drawbacks.

4.5.1 Value-at-Risk (VaR)

By far the most popular risk measure is called Value-at-Risk. Fixing some time horizon T it measures the amount of loss that is exceeded with probability $1 - \alpha$ within this time period.

Definition 18 (Value-at-Risk)**.** *Let* L *denote the loss of some risky asset over some time horizon* $[0, T]$. *The Value-at-Risk (VaR) at level* α *is defined as*

$$VaR_\alpha(L) = \inf\{x \in \mathbb{R} | \mathbb{P}(L > x) \leq 1 - \alpha\}. \tag{4.14}$$

In practice, α takes high values, e.g. 0.95, 0.99 or even 0.999, to take extreme events into account. The latter is necessary since these events may cause financial institutions to fail. As far as operational risk is concerned, an α level of 0.999 is the focus of interest.

The main disadvantage of using the VaR as risk measure is its lack of being subadditive and hence, not being a coherent risk measure. A simple example that underlines this statement can be found, e.g., in McNeil et al. [2005, Example 6.7]. The reason is mainly the non-consideration of the behaviour in the tail above the α level, i.e., the VaR takes the same value no matter if the remaining $1 - \alpha$ probability mass is located next to VaR_α or far away from it. This property keeps the usage of VaR facing criticism.

4.5.2 Expected Shortfall (ES)

On this account another risk measure has been gaining a lot of popularity during recent years. The so-called Expected Shortfall (also: Conditional Value-at-Risk or Tail Value-at-Risk) measures the mean amount of loss given that the loss exceeds VaR_α.

Definition 19 (Expected Shortfall)**.** *Let L denote the loss of some risky asset over some time horizon $[0, T]$. The Expected Shortfall of L at level α is defined as*

$$ES_\alpha(L) = \frac{1}{1-\alpha} \int_\alpha^1 VaR_\gamma(L) d\gamma$$

Since the upper limit of the integral in the definition of the ES is 1, this risk measure clearly does not suffer from ignoring the uppermost tail of the loss distribution. As a matter of fact, it can even be shown that the ES satisfies all the risk measure axioms, making it a coherent risk measure. The major drawback is the lack of being finite for all loss distributions (e.g., Cauchy or Power Law distributions with tail index $\alpha < 1$).

Because of the beneficial coherence property, the Basel Committee on Banking Supervision decided to replace VaR with ES in future internal model-based approaches for market risk, which was agreed on in the recent Fundamental Review of the Trading Book (FRTB), see Basel Committee on Banking Supervision [2013]. It is likely that other risk types will follow.

Chapter 5

Implementation of a Loss Distribution Approach

After having introduced a number of theoretical aspects to consider when fitting a statistical model, we now turn towards the applied part of this thesis. With the example of US businesses loss data, we develop a Loss Distribution Approach (LDA) incorporating the theoretical considerations from the previous part.

5.1 Data Description

The dataset at hand consists of operational losses incurred by some large businesses in the United States. Due to privacy issues not much more can be said about the data's origin. However, we stress that it is the same data that was used in Ainura Tursunalieva's PhD thesis (2012) at Monash University. Operational losses are listed on an aggregated monthly basis between May 1985 and June 2008, classified into seven different event types and nine different business lines as can be seen in Table 5.1. Note that this classification differs slightly from the one proposed by Basel II in Chapter 3.

As a result of the data being aggregated on a monthly level, it is neither possible to obtain the single loss event structure, nor can one map the losses uniquely to the Event Types/Business Line matrix. Moreover, we find the business line classification to be featuring several lines with very few loss observations (e.g., 19 observations in AC, 28 in RB), which is of small practical use when modeling each class individually. The event type segmentation yields clearly more even spread segments with a minimum of 51 observations for each. For this reason, we will consider a LDA with seven event types and ignore the

Table 5.1: Event Types/Business Lines classification

ET	Description
CPBP	Clients, products & business practices
DPA	Damage to physical assets
EPWS	Employment practices & workplace safety
EDPW	Execution, delivery & process management
EF	External fraud
IF	Internal fraud
OTHER	Remaining losses

BL	Description
CF	Corporate finance
TS	Trading & sales
RB	Retail banking
CB	Commercial banking
PS	Payment & Settlement
AC	Agency services
AM	Asset management
RG	Retail brokerage
OTHER	Remaining losses

business lines in the sequel. The data covers a time horizon of roughly 23 years in total. Noting that prices change over time, a correction for inflation is necessary to compare the data. If we refrain from adjusting the data, we may hardly assume independent and identically distributed losses which is the typical assumption in the LDA (however, there are publications relaxing this assumption, see e.g. Chavez-Demoulin et al. [2006]). Indeed, when examining the U.S. Consumer Price Index (CPI) for the relevant time period, taken from Bureau of Labor Statistics, United States Department of Labor [2016], in Figure 5.1, we find significant differences that need to be taken into account before performing data analysis. Thus, the inflation adjustment is as follows

$$L_t^{adj} = L_t \times \frac{CPI_{\text{June 2008}}}{CPI_t}, \tag{5.1}$$

where L_t^{adj} and L_t denote the inflation adjusted loss at time t and the original loss at time t, respectively. As can be seen in (5.1) the losses are adjusted to the consumer price level of June 2008.

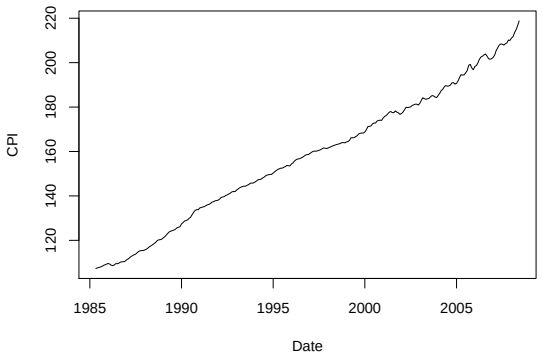

Figure 5.1: U.S. Consumer Price Index (1982-84 = 100)

5.2 Modeling the Severity

The first thing we tackle is the severity distribution. Usually, finding an appropriate severity model is regarded to be far more difficult when compared to modeling the frequency. Looking at the pie charts of loss severities in the respective event types in Figure 5.2, one may apprehend why this is the case. For almost all event types only very few losses cause a majority of the total loss amount within the respective segment. This heavy-tail behaviour can be seen as a common feature of operational risk data and a similar pie chart can be found, e.g., in Rootzén and Tajvidi [1997, Fig. 1]. However, what is striking is the fact that CPBP and EDPW exhibit even more extreme events than what is mentioned in Rootzén and Tajvidi [1997]. In particular we would like to highlight that one single CPBP loss event in July 1997 generates almost 75% of all CPBP losses within the studied time horizon. It casts doubt on whether usual extreme value theory models will be sufficient to obtain a reasonable fit. To facilitate readability of this text, only these two event types are considered in the sequel. Figure 5.3 shows the operational losses over time for these event

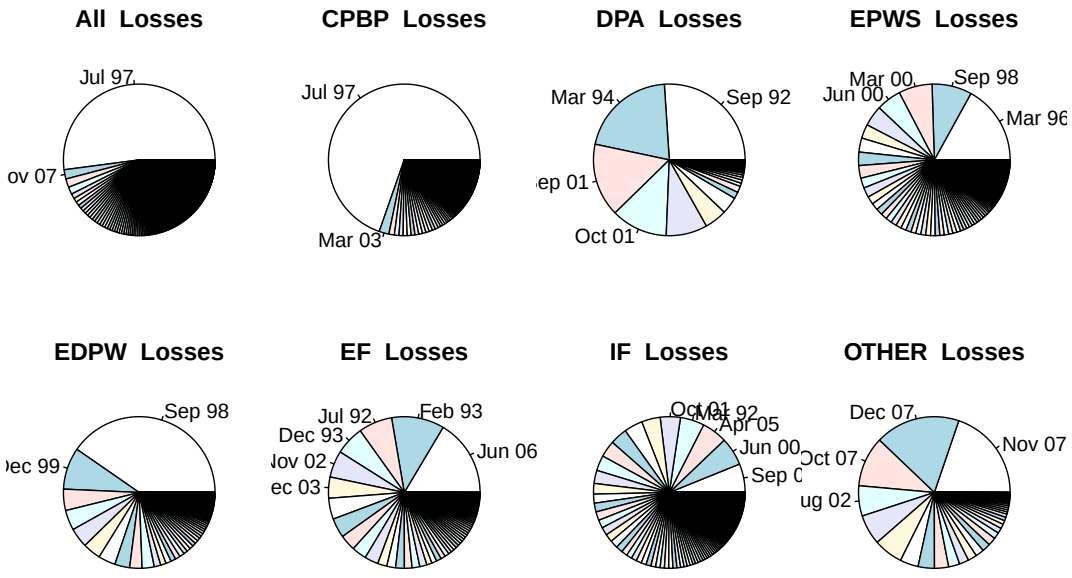

Figure 5.2: Pie charts of loss severities

types. The big difference in terms of severity between these event types becomes apparent. The *CPBP* event type happens to produce even more extreme losses than it is the case for *EDPW*. Naturally, one should cast a skeptical eye on an outlier on the scale of more than 400 billion USD in prices of June 2008. This amount is about the same magnitude as the

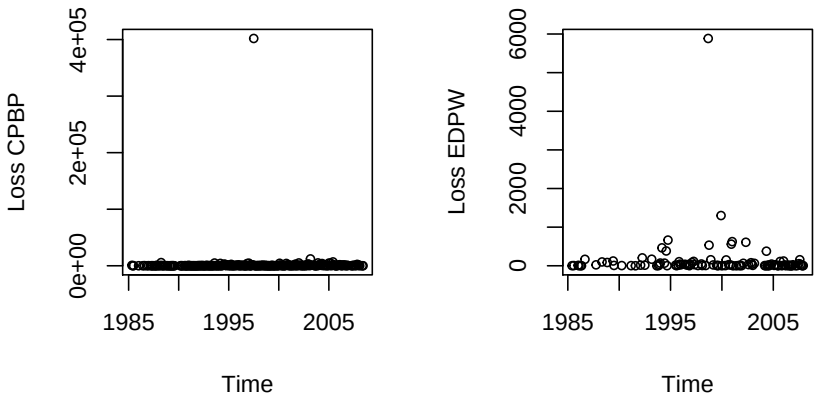

Figure 5.3: Operational losses over time

gross domestic product of Sweden. It cannot be entirely ruled out that this observation is caused by a typing error, because the information about this dataset is so scarce. In several applications eliminiating this data point would be a reasonable thing to do as it might affect the model estimation negatively. However, considering the large dependence of the final capital requirement on catastrophic events in the tail of the distribution, this could be a dramatic mistake and we refrain from doing so. In addition, experience has shown that these situations with one huge outlier occur more frequently than one might expect and we are going to present a new approach to deal with it.

In order to model the general data structure, typical heavy-tailed parametric distributions with non-negative support were fitted to the data initially. To be more precise, our analysis contained Lognormal, Weibull, Gamma and the Generalized Champernowne distribution. The obtained Q-Q plots are shown in the left-hand plots of Figure 5.4 with a scaled version of the same plots on the right.

We find that none of the univariate distributions can grasp any of the data structures accurately. Looking at the upper right plot, one observes a promising fit of the Lognormal and Generalised Champernowne distribution up to a threshold of around 700 million USD for event type *CPBP*, but they deteriorate significantly as soon as this threshold is exceeded. This can be explained by the fact that for the tail a tradeoff between the numerous regular exceedances and the outlier (see upper left plot) has to be found. The result is a tail that is actually too heavy for modeling the majority of the exceedances, but on the other hand too light to explain the outlier. Solving this quandary is the starting point of our future analysis.

But before this item is discussed in detail, we highlight that a different conclusion can be made from event type *EDPW*, despite its similiarity as far as the outlier is concerned. The lower plots in Figure 5.4 imply that again the lognormal distribution yields the best fit. In contrast to what we have observed for *CPBP*, even the outlier itself is reasonably well fitted. Nevertheless, it needs to be mentioned that the overall fit is still poor, clearly overestimating several quantiles between 300 and 3000. The poor results from the simple

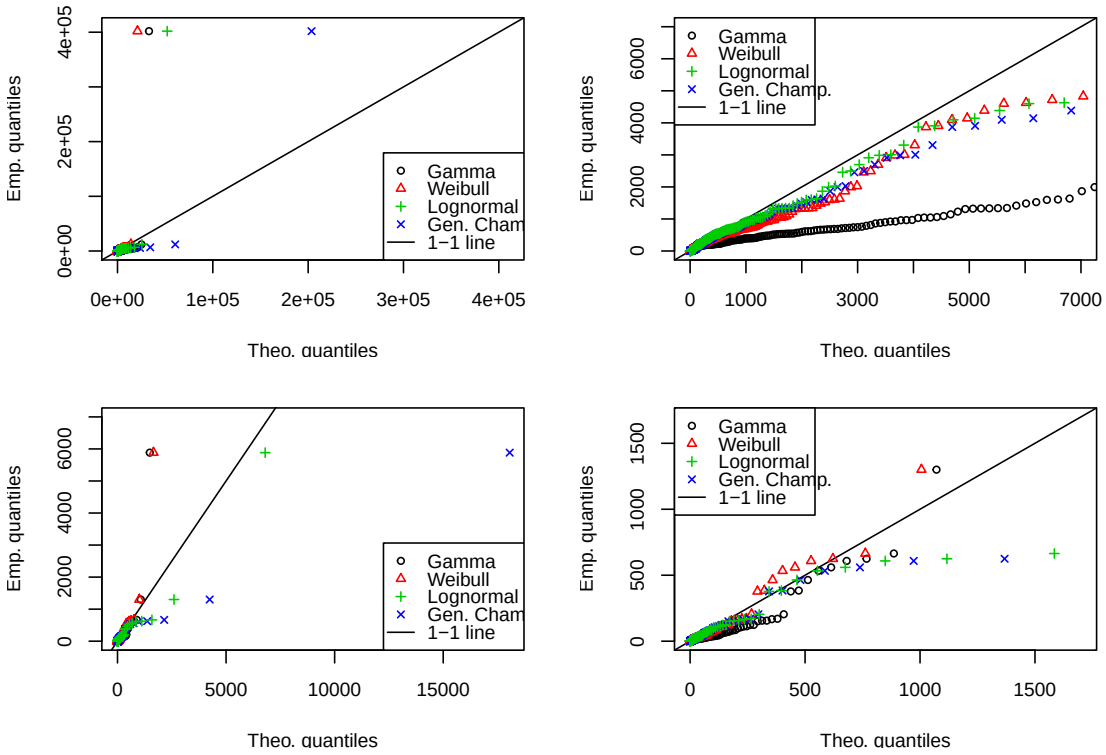

Figure 5.4: Full (left) and scaled version (right) of Q-Q-Plots for event types *CPBP* (upper row) and *EDPW* (lower row)

model stage motivate the use of more advanced tools and hence, a Peaks-over-Threshold approach is applied. It will also help us explaining why the outlier in the *CPBP* data worsens the tail fit, whereas the one in *EDPW* does not. The importance of threshold selection within the POT approach has been underlined several times before within this thesis and we make use of the tools presented in Section 4.3.1 to simplify the selection procedure. Particularly, we want to stress that it is a combination of all those tools which forms our decision.

Starting with the sample mean excess plot, a seemingly linear relationship is found

above a threshold of 4400 in the *CPBP* case, whereas for *EDPW* almost the entire plot looks linear, ignoring the oscillating bit between 400 and 600. This can be retraced in the upper plots of Figure 5.5, to which the least squares regression lines were added. Taking $u_{CPBP} = 4400$ and $u_{EDPW} = 20$ as initial guesses, the lower plots show the linear relationships above these respective thresholds. In these plots the dashed line marks the regression line obtained from removing the outlier and illustrates how these outliers affect the tail fits. While we observe a significant improvement for the fit of the non-outlier exceedances in the lower left figure, the inclusion of the outlier in *EDPW* hardly changes the regression line's slope. We draw the conclusion that the outlier in event type *EDPW* is more in accordance with the remaining data from a heavy-tailed distributional point of view than the one in *CPBP*. This elucidates the different behaviours described in Figure 5.4. The analysis of the corresponding sample median excess plots in Figure 5.6 raises

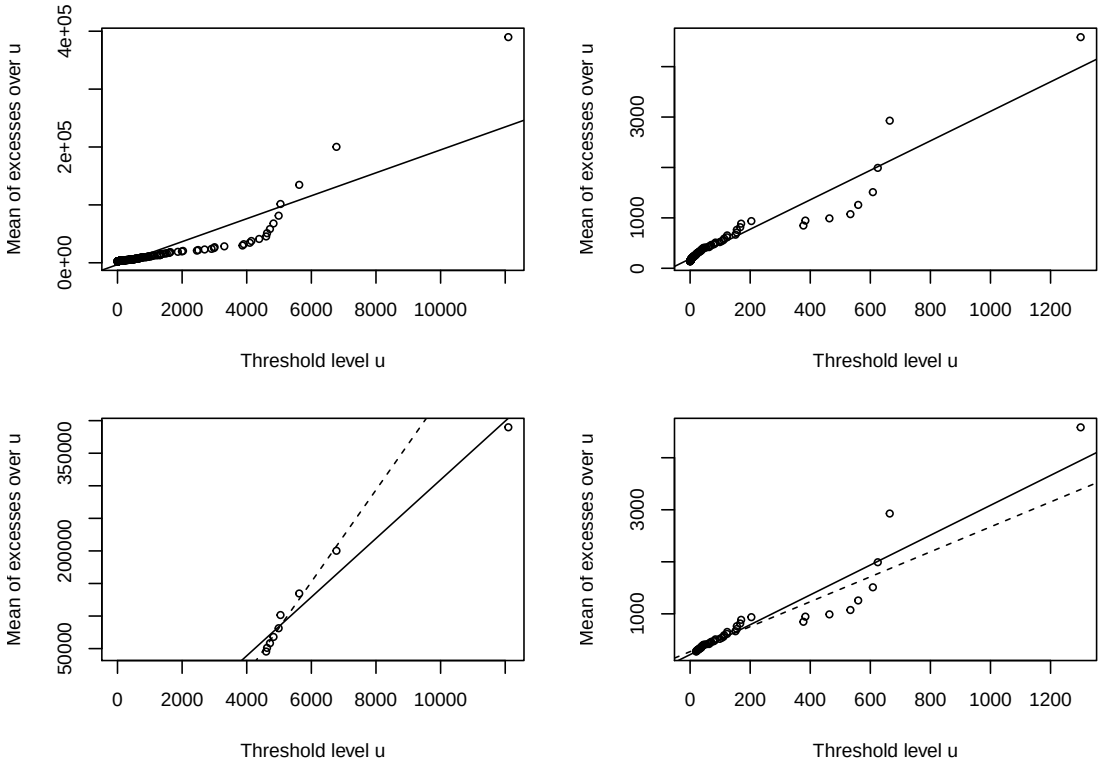

Figure 5.5: Sample mean excess plots for event types *CPBP* (upper left: all observations, lower left: observations above 4400) and *EDPW* (UR: all obs., LR: obs. above 20)
with least squares line containing the outlier (solid) and without the outlier (dashed)

the question whether it is even sufficient to model the entire *EDPW* severity by means of a Generalised Pareto distribution, i.e., $u_{EDPW} = 0$. As far as the threshold of *CPBP*

is concerned, we again find a linear relationship above 4400 if the effects of the outlier are omitted. However, we would like to stress that in contrast to its original motive, the sample median excess plot appeared to be less precise than the sample mean excess plot in a multitude of own simulations. Therefore, we see the obtained threshold suggestions more critically.

Having found some ideas about reasonable threshold values, we now proceed by fitting a GPD model to the respective exceedances and examine the resulting tail plots. The estimated parameters are listed in Table 5.2. Evidently, changing the threshold from 0 to 20 for *EDPW* has a statistically significant effect on the parameter estimates. The shape estimate decreases while the scale estimate increases. A similar effect is observed if the threshold is increased once more to a value of 25. All of the fitted models, except for $u_{EDPW} = 25$, are infinite mean models as $\xi > 1$.

> "This means that VaR may lead to ridiculously high capital charges in the infinite mean case ($\alpha < 1$). One could even argue that infinite mean models ought be banned from the operational risk modelers toolkit!", Nešlehová et al. [2006, p. 5].

In view of this quotation, infinite mean models are widely considered as dangerous, particularly if we are interested in very high quantiles such as the 99.9% quantile. For this reason, we prefer the choice $u_{EDPW} = 25$. As far as the shape parameter of the *CPBP* model is concerned, a model extension will solve this issue. But before this item is addressed in full length, a look at the tail plots in Figure 5.7 confirms once more how the huge outlier in *CPBP* is incompatible with the remaining observations, whereas in *EDPW* this is not the case. To make our models complete, we fit distributions to the body regions. Noting first that the total distribution function can be written as

$$\begin{aligned} F(x) &= \mathbb{P}(X \leq x) = \mathbb{P}(X \leq x | X \leq u)\mathbb{P}(X \leq u) + \mathbb{P}(X \leq x | X > u)\mathbb{P}(X > u) \\ &= p\mathbb{P}(X \leq x | X \leq u) + (1-p)\mathbb{P}(X - u \leq x - u | X - u > 0) \\ &= \begin{cases} p\frac{G(x)}{G(u)}, & \text{if } x \leq u, \\ p + (1-p)H(x-u), & \text{otherwise,} \end{cases} \end{aligned} \tag{5.2}$$

where $p = \mathbb{P}(X \leq u)$ denotes the probability of not exceeding the threshold and H is the estimated Generalised Pareto cdf. The former can be estimated by $\hat{p} = n_e/n$ with n_e and n being the number of exceedances and total number of observations for each event type, respectively. The only thing remaining is to find an estimate for the distribution function

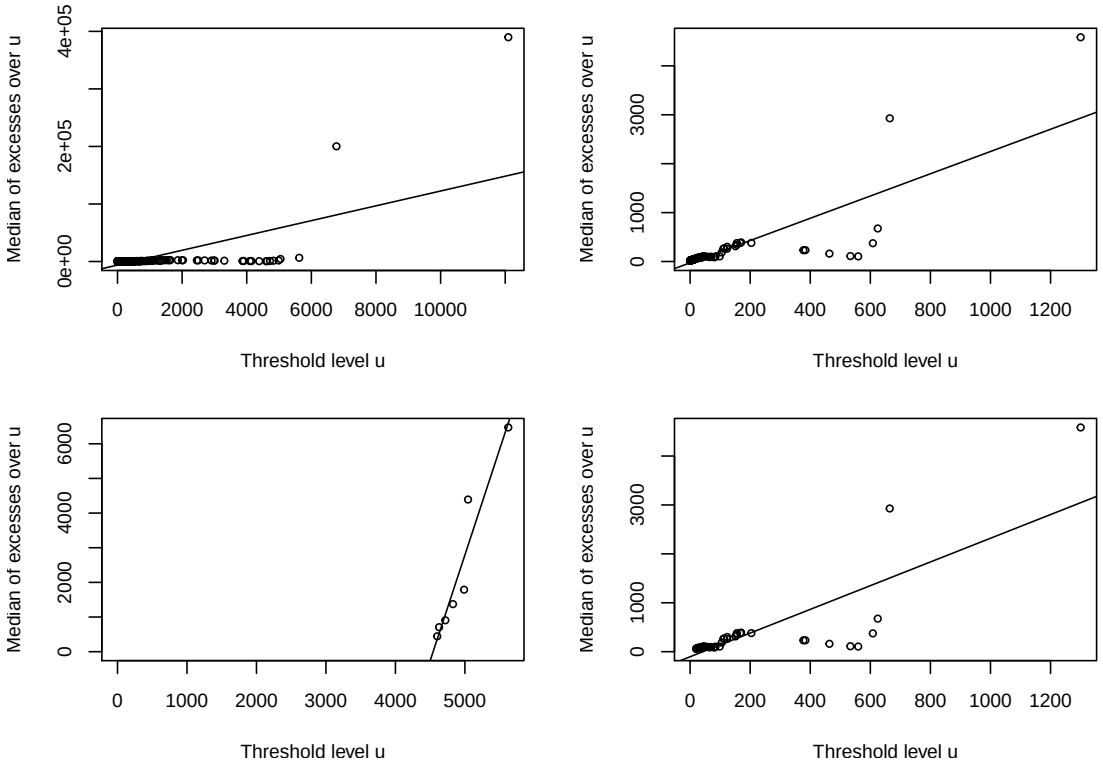

Figure 5.6: Sample median excess plots for event types *CPBP* (upper left: all observations, lower left: observations above 4400) and *EDPW* (UR: all obs., LR: obs. above 20) with least squares line containing the outlier

G which models the body of the data accurately, i.e., suits the observations below the threshold. Figure 5.8 show the Q-Q plots below the threshold if G is chosen as a lognormal and a Weibull distribution, respectively. These choices have proven to give the best fit to the corresponding event type. Indeed, the plots indicate an outstanding fit in the bottom part of the distribution. At the right end of the body one observes a slight underestimation of the quantiles, but that is less of an issue, since the 99.9% VaR is mostly determined by tail events rather than upper body events. The estimated model parameters can be found in Table 5.3. The Q-Q plots in Figure 5.8 and tail plots in Figure 5.7 give us an idea of whether the fitted models are sufficient to explain the data, but to get the whole picture, a formal statistical test needs to be applied. We decided to use the Anderson-Darling test in that matter, as it focuses on the tail of the distribution, which is what we are mainly interested in. The critical values are obtained from Monte Carlo simulation and the results for different levels of significance are found in Table 5.4. They all prove to be greater than

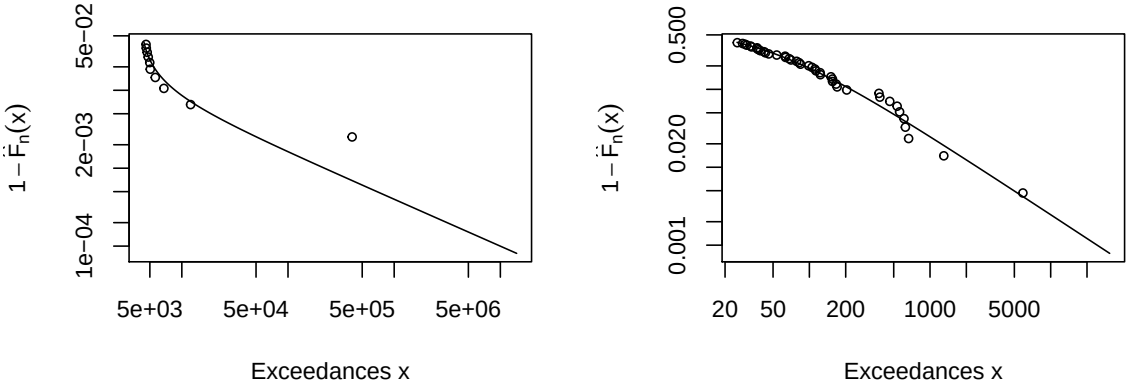

Figure 5.7: Tail plots for *CPBP* (left) and *EDPW* (right)

Table 5.2: Fitted GPD parameter estimates and standard errors

ET	Threshold	Parameter	ML estimate	Std error
CPBP	4400	Shape ξ	1.664	0.7362
		Scale β	737.0	445.3
EDPW	0	Shape ξ	1.483	0.2590
		Scale β	11.37	2.688
EDPW	20	Shape ξ	1.010	0.3231
		Scale β	48.87	15.46
EDPW	25	Shape ξ	0.9836	0.3163
		Scale β	65.70	20.86

the test statistics which are

$$A^2_{CPBP} = 1.21 \qquad \text{and} \qquad A^2_{EDPW} = 1.02.$$

Table 5.4: Simulated critical values for Anderson-Darling test for different levels of significance

ET	**90%**	**95%**	**99%**
CPBP	1.94	2.49	3.88
EDPW	1.91	2.47	3.80

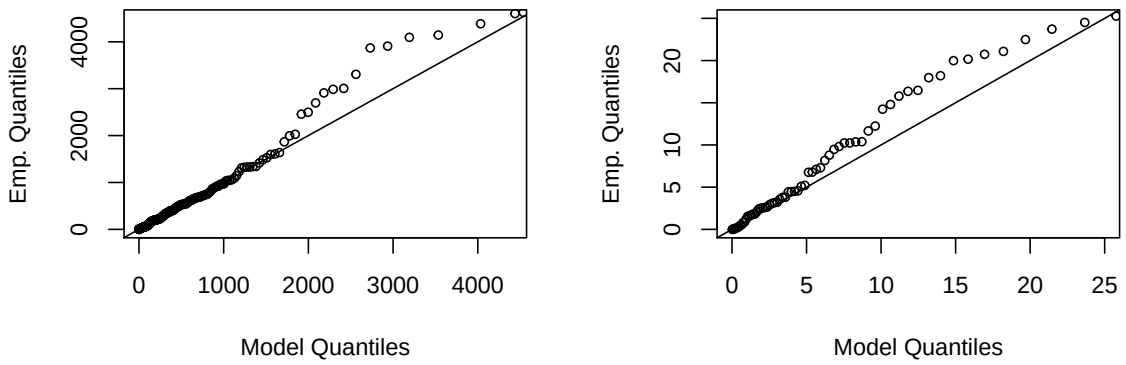

Figure 5.8: Q-Q plot derived from the total distribution in (5.2) for *CPBP* (left) and *EDPW* (right) with G being a lognormal and Weibull distribution, respectively

Table 5.3: Parameter estimates for event types *CPBP* and *EDPW* with G being lognormal and Weibull, respectively

ET	**Parameter**	**ML estimate**	**Std error**	**Profile LLH 95% CI**
CPBP	Location μ	5.135	0.1177	$(4.904, 5.365)$
	Scale σ	1.793	0.08326	$(1.642, 1.970)$
EDPW	Scale λ	6.900	1.067	$(5.221, 9.336)$
	Shape k	0.8445	0.08747	$(0.6872, 1.014)$

Modeling under the Presence of Huge Outliers

In this section, we want to introduce a new approach to handling huge outliers like the one in the event type *CPBP*, shown in Figure 5.2. Needless to say that *huge* refers to the relative relationship between the size of the outlier and the size of the remaining observations and has nothing to do with differences measured in absolute terms. As we have seen, e.g. in Figure 5.7, the existence of a huge outlier does not necessarily cause problems when fitting the tail. However, for *CPBP* this is certainly not the case. The outlier is not only poorly fitted, but its presence even deteriorates the fit for the remaining tail events. Moreover, a very high shape parameter is necessary to allow for the possibility of an event of this scope, resulting in an infinite mean model. The latter is critical in the context of high quantile estimation, as the quotation of Nešlehová et al. [2006] has shown. This brings up the question how this problem can be solved without contradicting the recognised extreme

value theory.

We remind ourselves how the shape and scale parameter in the POT model were obtained from a theoretical point of view. They both appear initially in the Extremal Types Theorem and the scale parameter needs to be transformed according to (4.9). Recalling that the Extremal Types Theorem assumes i.i.d. random variables, one can challenge whether this is a reasonable assumption. We want to give a motivation for a model in which the scale parameter switches between two different values.

In the spirit of Coles' chapter on extremes of non-stationary sequences, we allow the GPD scale parameter to take two different values. Mathematically, the model becomes

$$\beta(t) = \begin{pmatrix} 1 & S(t) \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 - \beta_1 \end{pmatrix},$$

where $S(t)$ is the indicator function taking a value of 1 if the observation at time t is an outlier and 0 otherwise. The idea of this small adjustment is to allow the outlier to stem from the same GPD with a different scaling, thus yielding more modeling flexibility but maintaining the possibility to estimate it reasonably well. Surely, this approach is not restricted to exactly one outlier, but can be applied even to a group of outliers. In the *CPBP* example, it is rather obvious which observation we want to assign a little extra model flexibility. The number of outliers needs to be considered when modeling the probability of the time series $(S(t))_{t\in\mathbb{Z}_+}$. We assume S to follow a homogeneous Markov chain and estimate the transition probabilites by the empirical probabilities.

Let the huge outlier be the j-th observation of the *CPBP* losses $x_1, \ldots, x_j, \ldots, x_n$. The likelihood function becomes

$$\mathcal{L}(\xi, \beta_1, \beta_2) = \left(\frac{1}{\beta_2} \left(1 + \xi \frac{x_j}{\beta_2} \right)^{-(\xi+1)/\xi} \right) \prod_{i \neq j} \frac{1}{\beta_1} \left(1 + \xi \frac{x_i}{\beta_1} \right)^{-(\xi+1)/\xi}$$

and as usual we maximise the loglikelihood function

$$\log \mathcal{L}(\xi, \beta_1, \beta_2) = -\log(\beta_2) - \frac{\xi+1}{\xi} \log \left(1 + \xi \frac{x_j}{\beta_2} \right) - (n-1) \log(\beta_1) - \frac{\xi+1}{\xi} \sum_{i \neq j} \log \left(1 + \xi \frac{x_i}{\beta_1} \right).$$

Table 5.5 lists the obtained parameter estimates as well as their corresponding 95% confidence interval using profile likelihood. What is striking is the low value of ξ giving us a finite mean model. Additionally, a big difference between β_1 and β_2 can be observed

with none of the 95% confidence intervals containing the other estimate. It is self-evident that due to the scarcity of exceedances, the confidence intervals are relatively wide. This holds in particular for β_2 which is determined by only one single observation (although, it is influenced by the estimate of ξ for which all observations are taken into account).

Table 5.5: Fitted GPD parameter estimates and standard errors

ET	**Threshold**	**Parameter**	**ML estimate**	**Profile LLH 95% CI**
CPBP	4400	ξ	0.4307	$(0.03501, 1.723)$
		β_1	853.2	$(361.4, 2186)$
		β_2	3.970×10^5	$(4.299 \times 10^4, 8.334 \times 10^6)$

Since from a statistical point of view, it is questionable to add another model parameter because of only one single observation, the consequences in terms of goodness-of-fit are considered. The tail plot in Figure 5.9 is examined. The red line shows the developing that is obtained from the scale switching GPD model. Compared to the regular fitted GPD (black line), it is clearly more flexible and fits both, the regular exceedances and the outlier, very accurately. In contrast, implementing the same idea but switching the shape turns out to be less helpful, as it yields a worse fit and an infinite mean model. In short, the scale-switching mixed GPD features an excellent fit of the tail and there is hope that even formal tests prove it to be superior than regular GPD. The tests which are used to

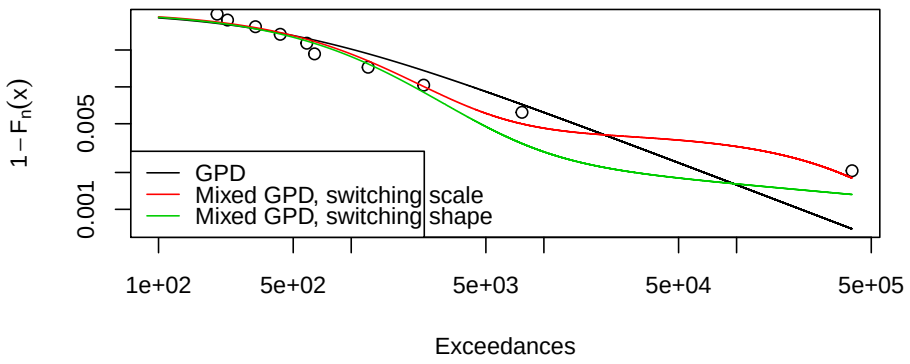

Figure 5.9: Tail plot for the mixed GPD model for event type *CPBP*

assess the quality of the model are the Likelihood Ratio test and the Anderson-Darling

test. The former test aims at answering the question whether the extra gain is sufficient to justify the extension of the regular GPD model, whereas the AD test considers the fit from an absolute rather than a relative perspective. The test statistics

$$LR = 2(\log \mathcal{L}_{mixGPD} - \log \mathcal{L}_{GPD}) = 2(-87.938 + 92.665) = 9.4552$$

is considerably greater than the critical value $\chi^2_1(0.95) = 3.84$, such that the null hypothesis $H_0 : \beta_2 - \beta_1 = 0$ of the Likelihood Ratio test can be rejected.

The Anderson-Darling test can be applied to either the whole data or the exceedances only. Depending on our choice, the critical value needs to be changed. Since the scale-switching mixed GPD is clearly not a standard distribution, the critical values are obtained from Monte Carlo simulation. As the AD test already puts a high weight on the tails, the entire *CPBP* dataset is used and compared to the critical value which one gets by simulating from the Lognormal-Mixed-GPD-model. These critical values are shown in Table 5.6.

Table 5.6: Critical values for Anderson-Darling test for different levels of significance

90%	95%	99%
106.67	112.74	124.12

The calculated test statistics

$$AD = -N - S = -242 + 317.04 = 75.04$$

lies below all of the listed critical values, i.e. the fitted model can not be rejected at any reasonable level of significance. Thus, we have found a model that seems to cope reasonably with outliers and yields a sufficient overall fit. These properties need to be satisfied in order to avoid awful simulation results.

5.3 Modeling the Frequency

Although the main focus of the majority of literature, covering the quantification of operational risk, is on the severity distribution, a consistent frequency model should not be neglected. In this chapter, we present a new way of modeling the frequency when the data is reported on a monthly basis.

It is common practice to choose a Poisson process when it comes to the loss frequencies. This does not only yield a reasonably good fit, but it is also convenient due to the model and estimation simplicity. More ambitious practitioners accept the more complex Negative Binomial process in order to gain a better model fit as well as getting rid of the "mean equals variance" assumption. There are also other processes possible, but those are rarely seen in both, theory and practice.

Unlike the common reporting of each loss individually, the special feature with our data is the monthly loss reporting. It is not difficult to realise why neither a Poisson nor a Negative Binomial process are suitable choices for our purpose. If we consider a 12-months horizon, the number of loss events is an integer between 0 and 12 and therefore bounded from above. Both, the Poisson distribution and the Negative Binomial distribution are not bounded from above, which will put positive probability mass to $13, 14, \ldots$ loss months. Unless this mass is negligible (e.g., because the Poisson rate λ is close to 0), this forces us to think about model alternatives.

To understand the structure of the data, we plot the number of months with positive losses for each year and check initially, if there is any hint for time dependence. The time series plots in Figure 5.10 support the notion that dependence over time is in fact an issue. Particularly, the *CPBP* time series shows a strong persistence in one state. Though the *EDPW* loss frequencies are less persistent, huge jumps from one year to another are very rare, too. For this reason, the classical i.i.d. assumption is invalid and more sophisticated tools are required. In this section we will be discussing two different ways for solving this time dependence issue. In order to make formal tests with modeling-independent data, only the information from 1985-2000 is taken for the modeling part. The remaining data are used for testing the model correctness.

Markov Chain Approach

From Figure 5.10 it can not be justified to include some kind of a trend or seasonal component in the model, but the number of loss months is clearly related to the number of loss months one year before. Hence, the first approach is based on Markov chains and allows for state-dependent transition probabilities. The number of data points in Figure 5.10, however, is small, making it impossible to estimate the transition matrix $\boldsymbol{P}$ on the state space $\{0, 1, \ldots, 12\}$.

Transforming the problem back to a monthly rather than a yearly scale not only yields a higher number of observations, but also provides some degrees of freedom as far as the

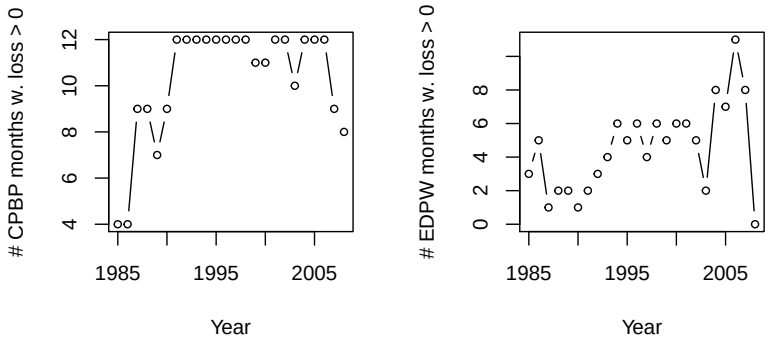

Figure 5.10: Number of annual loss months for event types *CPBP* (left) and *EDPW* (right)

state space choice is concerned. The simplest idea would be to consider $S = \{0, 1\}$, where the Markov chain $(X_k)_{k\in\mathbb{N}}$ takes state 1 if an operational loss has occurred and 0 otherwise. Figure 5.11 shows the resulting state developing for both event types. While the *CPBP* Markov chain (left plot) does not look time-homogeneous at all, an evenly spread plot is found for *EDPW*. The former observation raises the question whether a Markov chain is a reasonable choice for modeling the *CPBP* OpRisk loss frequencies.

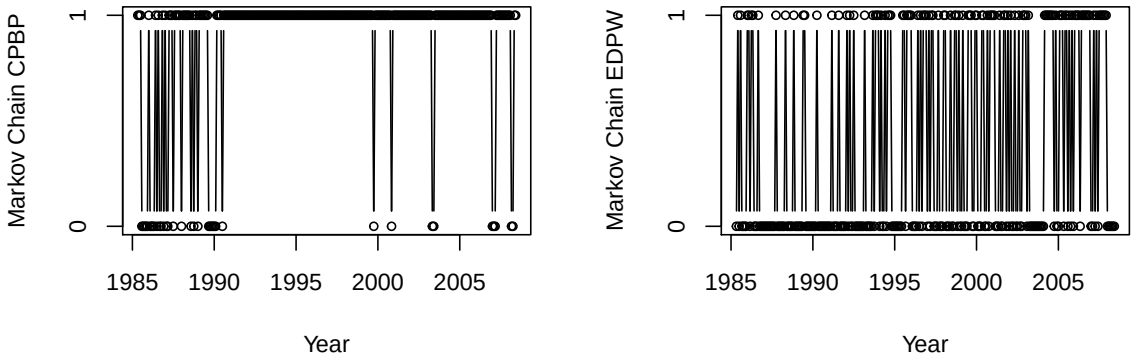

Figure 5.11: Indicator function of monthly reported operational losses for event types *CPBP* (left) and *EDPW* (right)

Noting the structural change in event type *CPBP* in 1991, one might ask if some event (e.g., internal efforts to report operational losses) caused this increase of loss months. With regard to the few months after 1991, in which no losses were reported, the inclusion

of the first six years into the Markov chain estimation procedure would clearly lead to an underestimation of the probability of a loss month in 2008. Therefore, it is decided to remove the first ten years from the estimation part.

Given this adjustment, the estimated transition matrices are found to be

$$\widehat{\boldsymbol{P}}_{CPBP} = \begin{pmatrix} 0.4828 & 0.5172 \\ 0.0949 & 0.9051 \end{pmatrix} \qquad \text{and} \qquad \widehat{\boldsymbol{P}}_{EDPW} = \begin{pmatrix} 0.6953 & 0.3047 \\ 0.6441 & 0.3559 \end{pmatrix} \tag{5.3}$$

and to assess the model quality, the null hypothesis $H_0 : \boldsymbol{P} = \widehat{\boldsymbol{P}}$ is tested by means of the χ^2 test with test statistics

$$Q = \frac{(N_{00} - n_0\widehat{p}_{00})^2}{n_0\widehat{p}_{00}} + \frac{(N_{01} - n_0\widehat{p}_{01})^2}{n_0\widehat{p}_{01}} + \frac{(N_{10} - n_1\widehat{p}_{10})^2}{n_1\widehat{p}_{10}} + \frac{(N_{11} - n_1\widehat{p}_{11})^2}{n_1\widehat{p}_{11}},$$

where N_{ij} refers to the number of transitions from state i to j in the testing dataset. The values of $\widehat{p}_{ij}$ are given by the entries of the transition matrix under H_0. n_0 and n_1 are the frequencies of the chain being in state 0 and 1, respectively. The very last data point is ignored for the calculation of n_0 and n_1 as there is no consecutive data point available. The testing frequencies

$$\widehat{\boldsymbol{N}}_{CPBP} = \begin{pmatrix} 4 & 3 \\ 3 & 79 \end{pmatrix} \qquad \text{and} \qquad \widehat{\boldsymbol{N}}_{EDPW} = \begin{pmatrix} 23 & 19 \\ 20 & 27 \end{pmatrix}$$

yield $Q_{CPBP} = 3.47$ and $Q_{EDPW} = 14.04$. As under H_0 the test statistics follows a χ^2 distribution with 3 degrees of freedom, the critical value at a 95% confidence level is 7.81, such that H_0 can be rejected for *EDPW*, but not for *CPBP*. In view of Figure 5.11, however, it can be doubted that a homogeneous Markov chain is actually sufficient to model the *CPBP* loss frequency.

Although the MC model fails to be helpful for this particular dataset, it might be reasonable in other situations. One could even go one step further by extending the state space to $S = \{0, 1, 2\}$ where states 1 and 2 are attained if the monthly loss falls below or exceeds some threshold u, respectively. Thereby, the strict separation between loss frequency and loss severity is set aside. Letting u be equal to the threshold u' used for modeling the loss excesses, the simulation process is simplified. Note that even the concept of switching GPD parameters using a hidden MC can be easily incorporated into this model.

Heterogeneous Bernoulli Approach

For the second modeling approach the Markov assumption is rejected. In place of this assumption, we let the former Markov chain $(X_k)_{k\in\mathbb{N}}$ be a sequence of Bernoulli distributed random variables with time-dependent parameters $(p_k)_{k\in\mathbb{N}}$. Since Figure 5.11 does not give any evidence what kind of mathematical relationship can be imposed on p (e.g., linear trend or cyclical fluctuation), a concept from time series analysis comes into play. The recursive least squares estimation (see, e.g., Jakobsson [2013, pp. 256-259]) is a well-known approach for tracking dynamic systems. It says that a time-dependent estimate for p_k can be obtained from

$$\hat{p}_k = \arg\min_p \sum_{j=1}^{k} \beta_j (x_j - p)^2,$$

where x_j is the realisation of the rv X_j and β_j is the so-called forgetting factor. A typical choice for β is given by

$$\beta_j = \lambda \beta_{j+1}$$

with $\beta_k = 1$ and one tuning parameter λ that usually takes values slightly less than 1. The lower the value of λ is chosen, the more relative weight is put on the most recent observations in the minimisation problem, thus leading to a quickly adapting estimate of p_k. Solving the first order condition yields

$$\hat{p}_k = \frac{\sum_{j=1}^{k} \beta_j x_j}{\sum_{j=1}^{k} \beta_j}.$$

In particular, assuming equal weights yields the ML estimate $\hat{p}_{ML} = k^{-1} \sum_{i=1}^{k} x_i$. To avoid the arbitrariness of choosing a suitable value for λ, an optimisation problem given by

$$\hat{\lambda} = \arg\min_\lambda \sum_{j=J}^{k-1} (x_{j+1} - \hat{p}_j(\lambda))^2$$

is proposed, where J is a small integer that should prevent too small estimates of λ resulting from the need of a speedy adjustment in the first months. One year of adaption phase will be allowed, so $J = 12$. Figure 5.12 plots the values of this sum of quadratic deviations for different values of λ. The minima are found to be $\hat{\lambda}_{CPBP} = 0.839$ and $\hat{\lambda}_{EDPW} = 0.961$.

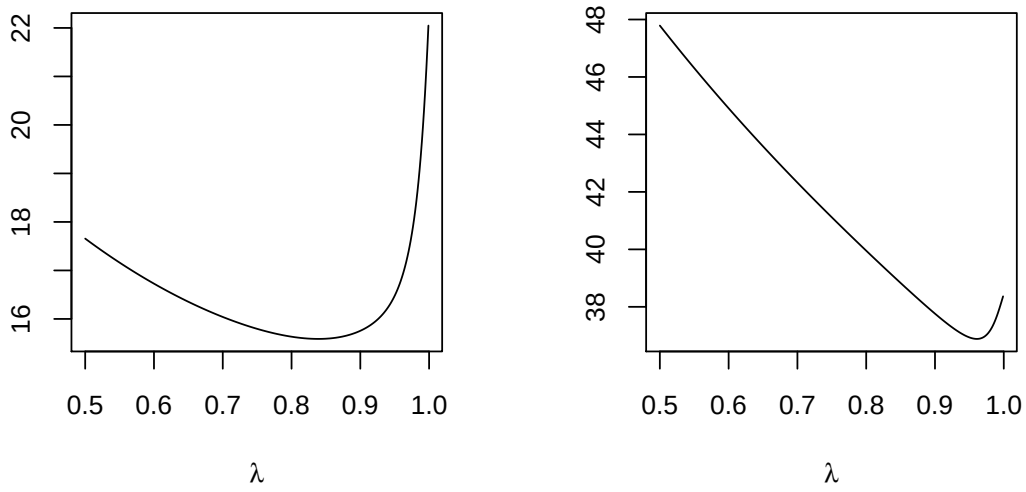

Figure 5.12: Sum of quadratic deviations for different choices of the tuning parameter λ for event type *CPBP* (left) and *EDPW* (right)

The developing of the time series $(\hat{p}_k^{CPBP})_{k\in\mathbb{Z}_+}$ and $(\hat{p}_k^{EDPW})_{k\in\mathbb{Z}_+}$ are displayed in Figure 5.13. Due to $\hat{\lambda}_{CPBP} < \hat{\lambda}_{EDPW}$ the CPBP loss probability is more sensitive to new data and hence, features a higher volatility.

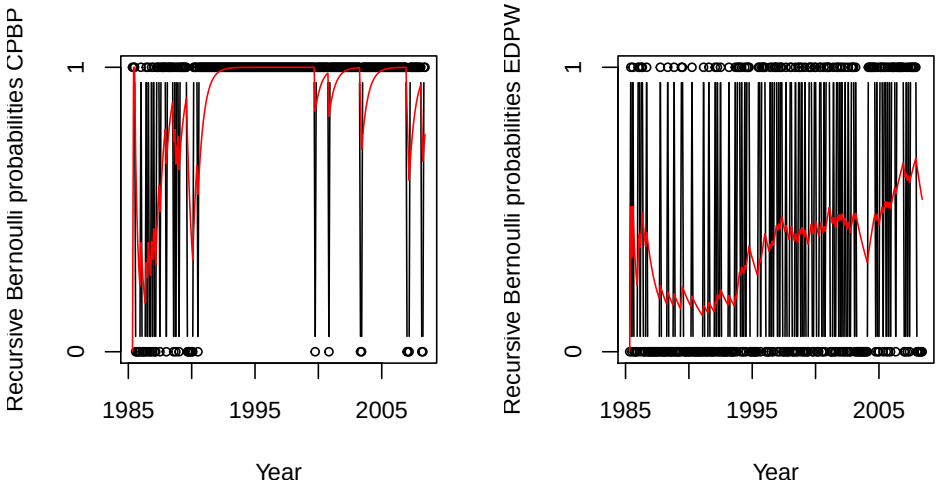

Figure 5.13: Recursive least squares Bernoulli probabilties (red) for event types *CPBP* (left) and *EDPW* (right)

In order to check the correctness of these models, a formal statistical test needs to be run. For this purpose, the data are assumed to stem from a heterogeneous Bernoulli

distribution with parameter time series (p_0). A test of the null hypothesis $H_0 : p_0 = \hat{p}$ is introduced by deriving the distribution of the test statistics

$$T = \sum_{j=J}^{k} (x_{j+1} - p_j)^2.$$

The distribution can be either found by simulation or one uses a normal approximation by applying Lyapunov's version of the Central Limit Theorem (see Billingsley [1995, p. 362]). The latter differs from the regular CLT by allowing the random variables to follow different distributions. Under H_0 it can be shown that

$$\mathbb{E}[T|p_0 = \hat{p}] = \sum_{j=J}^{k} \hat{p}_j - \hat{p}_j^2 \qquad \text{and} \qquad Var[T|p_0 = \hat{p}] = \sum_{j=J}^{k} -4\hat{p}_j^4 + 8\hat{p}_j^3 - 5\hat{p}_j^2 + \hat{p}_j$$

and by means of Lyapunov's CLT it follows that approximately, it holds

$$T \sim \mathcal{N}\left(\sum_{j=J}^{k} \hat{p}_j - \hat{p}_j^2, \sum_{j=J}^{k} -4\hat{p}_j^4 + 8\hat{p}_j^3 - 5\hat{p}_j^2 + \hat{p}_j \right).$$

The resulting test statistics do not exceed the critical value at a 95% level of significance, as can be seen in Table 5.7. Thus, the null hypothesis can not be rejected. The slightly differing values of $T_{0.95}^{Sim}$ and $T_{0.95}^{CLT}$ are mainly due to the imprecision of the normal approximation. The simulated critical value is preferable when evaluating the goodness-of-fit.

Table 5.7: Goodness-of-Fit test heterogeneous Bernoulli

Event type	$\boldsymbol{T}$	$\boldsymbol{T_{0.95}^{Sim}}$	$\boldsymbol{T_{0.95}^{CLT}}$
CPBP	6.46	6.55	6.80
EDPW	19.72	20.05	20.37

5.4 Simulations

So far, we have fitted models to the severity and the frequency of losses in *CPBP* and *EDPW* individually. In this section we aim at combining these models in order to determine

the 99.9% VaR of the loss distribution on a 12-months horizon. Formally, this can be described as finding the distributions of

$$L^{CPBP} = \sum_{i=1}^{12} L_i^{CPBP} \qquad \text{and} \qquad L^{EDPW} = \sum_{i=1}^{12} L_i^{EDPW},$$

where L_i^{CPBP} and L_i^{EDPW} denote the loss in month i for *CPBP* and *EDPW*, respectively. How are these losses connected to loss severity S and loss frequency X? Noting that X_i describes the indicator function of observing a loss in month i, we find the monthly losses to be simply the product of S and X, i.e.

$$L_i = S_i X_i.$$

Due to the models for S and X being rather complicated, there is no simple analytical solution for the distribution of L^{CPBP} and L^{EDPW}. In fact, this is even the case if simpler models are chosen. Hence, we make use of numerical simulations. The 99.9% Value-at-Risk can then be estimated from the empirical 99.9% quantile of the simulated realisations of L. The obtained values are shown in Table 5.8.

Table 5.8: Simulated 99% VaRs and bootstrapped 95% confidence intervals (in brackets) in million USD. 1 Million 12-months simulations were used.

CPBP	**EDPW**
$3.611(3.605, 3.618) \times 10^6$	$1.208(1.148, 1.279) \times 10^5$

5.5 Modeling the Dependence

Having found estimates for the risk measures in both event types, it needs to be asked how the risk of the entire portfolio of operational losses follows from these calculations. If, for some reason, extreme losses in *CPBP* coincide with small losses in *EDPW* and vice versa, the total risk is understandably lower than if extreme events happen at the same time in both event types. The Basel II regulations address this aggregational problem by suggesting banks to sum up their risk measures:

> "Risk measures for different operational risk estimates must be added for purposes of calculating the regulatory minimum capital requirement. However,

> the bank may be permitted to use internally determined correlations in operational risk losses across individual operational risk estimates, provided it can demonstrate to the satisfaction of the national supervisor that its systems for determining correlations are sound, implemented with integrity, and take into account the uncertainty surrounding any such correlation estimates (particularly in periods of stress). The bank must validate its correlation assumptions using appropriate quantitative and qualitative techniques.", Basel Committee on Banking Supervision [2006, p. 152].

There is a certain misbelief about the concept of adding all VaRs. Without any doubt, this method is based on the assumption of perfect dependence, i.e., for all observation pairs (L_{CPBP}, L_{EDPW}) there exists a $0 < p < 1$ such that L_{CPBP} is the p-quantile of the $CPBP$ marginal distribution and L_{EDPW} the p-quantile of the $EDPW$ marginal distribution. Broadly speaking, high losses in $CPBP$ coincide with high losses in $EDPW$ and similarly for low losses. However, perfect dependence does not imply the sum of VaRs to be the upper bound when comparing with other dependence structures. This is a consequence of the Value-at-Risk not being subadditive. With this in mind, one is inclined to ask if there is a general way to assess the upper bound. Embrechts and Puccetti [2006] address this question and develop a sophisticated numerical procedure to obtain upper bounds. The entire problem seems to be a non-trivial one, featuring no general solution.

The quotation of the Basel Committee contains but another important passage. Internally determined correlations are permitted under certain constraints regarding the accurateness and conservatism of such a model. In this section we develop a model which takes the correlation between $CPBP$ and $EDPW$ losses into account. Our aim is to compare the resulting risk measure with the sum of the event type based risk measures.

What can be said about the correlation between $CPBP$ and $EDPW$ losses? There is clearly no obvious explanation why these losses should be strongly correlated. But on the other hand simply assuming independence for this reason does not satisfy the requirements imposed by Basel II. Figure 5.14 shows a scatterplot of all 23 yearly losses in 1985-2007 on a log scale. It becomes apparent that there is in fact not much dependence between losses in the respective event types, although a weak positive relationship can be spotted.

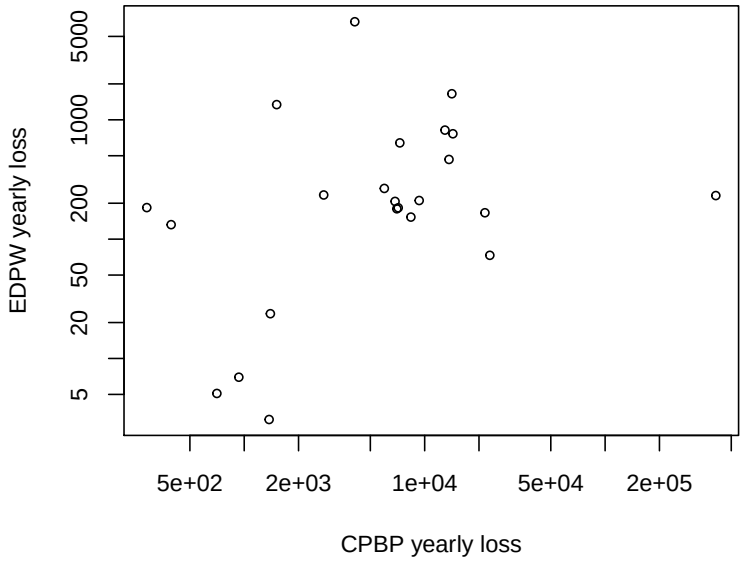

Figure 5.14: Scatterplot of yearly losses in event types *CPBP* and *EDPW* from 1985 to 2007

These observations are in accordance with the results in Table 5.9, listing the estimated Pearson's correlation coefficient ρ, Kendall's τ and Spearman's ρ_S for the underlying dataset. The latter two dependence measures take values that were to be expected after our former data analysis. The small negative value of Pearson's ρ is not a contradiction to Figure 5.14 because of the log scale axes and the invariance of Pearson's ρ under monotone transformations. Given the rather small values of dependence, it suggests itself to test for independence. The confidence intervals resulting from bootstrapping indicate that independence can not be rejected for this dataset.

Table 5.9: Estimated measures of dependence with 95% confidence interval obtained by bootstrap methods

Pearson's ρ	**Kendall's τ**	**Spearman's ρ_S**
-0.066 (-0.203,0.389)	0.233 (-0.089, 0.496)	0.367 (-0.117,0.686)

In order to benefit from the univariate analyses in Section 5.2 and 5.3, a copula approach is developed. This raises the question what copula family suits the data. Di Clemente and Romano [2004], one of the more popular papers using copulas in the context of OpRisk,

suggests to use Student's t-copula. The authors argue that it is more flexible due to the higher number of parameters and contains the famous Gaussian copula as a special case. This argumentation could be countered by stating that a higher number of parameters also makes the estimation procedure more cumbersome and does not necessarily give better results. Furthermore, we point out that Student's t-copula only allows elliptical dependence structures and, in addition, the use of the Gaussian copula has proven to be a model which needs to be faced with much caution in risk management due to the tail independence.

Therefore, this decision is delayed and the performance of different copula families is examined. Given the complexity of the marginal distributions, a Full-Maximum-Likelihood approach does not work in this context and hence, a two-stage IFM method is used to estimate the copula parameters. Note the difference to the Canonical Maximum Likelihood approach in Di Clemente and Romano [2004, p. 11].

Besides the proposed Student's t copula, three copulas (Frank, Gumbel and Clayton) of the Archimedean family are fitted to the data. The dependence estimates in Table 5.10 are difficult to compare, due to the different parametrisations of each copula type. Given the independent cases have parameter values zero (t and Clayton) and one (Frank and Gumbel), the small positive dependence is apparent in all estimates. Moreover, it can be concluded that the additional parameter in the t copula does not result in a superior likelihood, when compared to the other copula types. It is in fact the Clayton copula which minimises the Akaike Information Criterion (AIC) and the result of several Goodness-of-Fit tests in Table 5.11 does not raise an objection against this choice. A word of caution needs to be added when using a Clayton copula, as its upper tail dependence is zero. As a result, it assumes extreme operational losses in *CPBP* and *EDPW* to be independent from each other. In some contexts this might be a critical assumption. Here, this can be justified as there is no obvious explanation why tail events in *CPBP* and *EDPW* need to be dependent and Figure 5.14 supports this notion.

Table 5.10: Copula estimation results based on IFM (Standard error in brackets)

	Student's t	**Frank**	**Gumbel**	**Clayton**
Dependence	0.32 (0.16)	2.43 (1.16)	1.23 (0.17)	0.44 (0.22)
Degrees of freedom	13.6 (58.1)	–	–	–
Max. loglikelihood	2.26	2.39	1.25	3.32
AIC	-0.51	-2.78	-0.51	-4.64

Table 5.11: p-Values Goodness-of-Fit Test

	S_n	$S_n^{(B)}$	$S_n^{(C)}$	A_n
Clayton(0.44)	0.839	0.594	0.735	0.181

5.6 Computing the Capital Charge

The aggregated 99.9% Value-at-Risk of losses from *CPBP* and *EDPW* is given by the 99.9% quantile of the sum $L_{CPBP} + L_{EDPW}$. Formally, this can be expressed by

$$VaR_{0.999}(L_1 + L_2) = \inf\left\{x \in \mathbb{R} | \int_0^\infty \int_0^\infty \mathbb{I}(L_1 + L_2 > x) dC(F_1(L_1), F_2(L_2)) \leq 1 - \alpha\right\},$$

with F_1 and F_2 being the cdfs of *CPBP* losses and *EDPW* losses, respectively. Since both loss distributions can not be expressed by simple analytical terms, the Value-at-Risk estimate is found by means of simulation and empirical quantile calculation.

The expected loss EL is defined as the expected value of the aggregated lossed. In mathematical terms

$$EL = \int_0^\infty \int_0^\infty [L_1 + L_2]\, dC(F_1(L_1), F_2(L_2)),$$

which again can not be calculated explicitely, but needs to be estimated by the mean of a large sample of loss simulations of $L_{CPBP} + L_{EDPW}$.

The estimation results of Value-at-Risk, expected loss and minimum capital charge are listed in Table 5.12. For comparison, even the Markov chain loss frequency model was considered in this table, although there is some evidence that its use is less reasonable. It is apparent that all models seem to produce similar capital requirements. In particular, modeling the dependence does not entail a significant capital reduction, as one could have expected. Quite the contrary, there is even one case in which complete dependence produces a lower capital charge than the one obtained by using a fitted copula. However, this difference might be due to estimation imprecision.

While all VaR estimates are roughly of the same dimension, there is an increase of about 20% in EL when the Markov chain approach is used for *CPBP*. A detailed explanation for this difference as well as for other findings are discussed in the following chapter.

Table 5.12: Estimated VaR, EL and Capital Charges (in billion USD)

Freq CPBP	**Freq EDPW**	**Copula**	$\boldsymbol{VaR}_{0.999}$	**EL**	**Capital**
RB	MC	Perfect dep.	3,724	37.14	3,686
RB	MC	Clayton(0.44)	3,634	36.65	3,598
RB	RB	Perfect dep.	3,732	36.30	3,696
RB	RB	Clayton(0.37)	3,778	35.73	3,742
MC	MC	Perfect dep.	3,819	42.34	3,777
MC	MC	Clayton(0.20)	3,730	42.97	3.687
MC	RB	Perfect dep.	3,827	42.34	3,785
MC	RB	Clayton(0.20)	3,709	41.56	3,668

Chapter 6

Discussion

The results in Table 5.12 provide a good overview for both, data characteristics and the impact of model choice. This becomes apparent in the similar values of the VaR and the different values for EL when the loss frequency in *CPBP* is modeled by a Markov chain. The negligible effect of modeling the dependence is another result that originates from data characteristics.

To be more precise, we start by taking a look at the capital charge. With regard to equation (3.1) and Figure 3.1 this value is almost entirely determined by the 99.9% VaR, which exceeds the EL many times over in heavy-tailed distributions. The 99.9% VaR, however, mainly consists of the losses in event type *CPBP*, as they have proven to be clearly larger in the tail than *EDPW* losses. Considering the *CPBP* Markov chain transition matrix estimation result in (5.3) and the left plot in Figure 5.13, the probability of observing a *CPBP* loss month is quite large at the end of the time series in 2008. Thus, it is very likely that tail events in the simulated 12-months loss distribution feature also 12 months of losses. The choice of frequency distribution hardly makes a difference in that matter.

On the other hand, the frequency model affects the expected loss in the sense that the Markov chain approach for *CPBP* losses generates larger values for EL than the Recursive Bernoulli approach. Again, (5.3) and Figure 5.13 give a simple explanation for this behaviour. While the Markov model will almost exclusively produce 12 months of losses, thr Recursive Bernoulli model starts at a much lower probability of 0.76. Hence, the number of loss months will on average be lower than in the Markov model. Note the dependence on the initial time point and the probability at this date. Figure 6.1 emphasises this argument by showing the different histograms of loss months for both models obtained by simulation.

The heavier left tail of the RB model stands out.

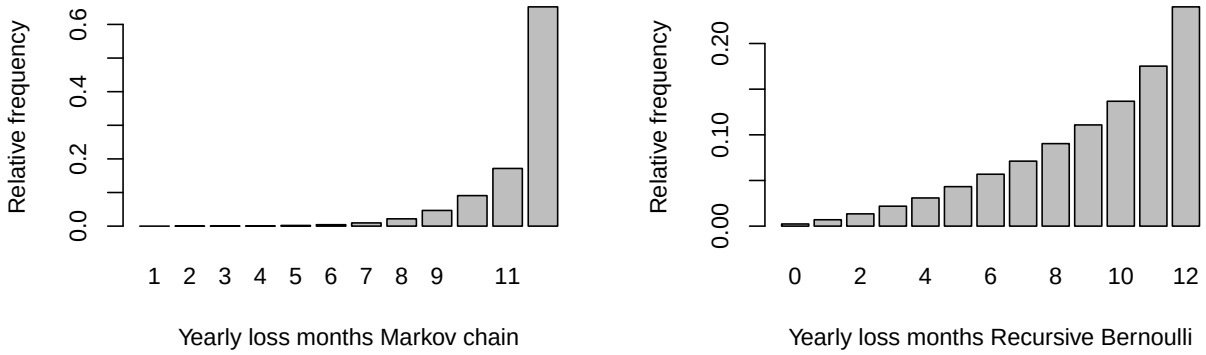

Figure 6.1: Histogram of yearly *CPBP* loss months starting in June 2008 when using a Markov chain (left) and a Recursive Bernoulli approach (right)

Another interesting finding from Table 5.12 is the small difference resulting from modeling the dependence. In fact, this outcome is hardly surprising given the disparity in scale of both event types. Even in a favorable scenario of perfect negative dependence (i.e., high losses in *CPBP* coincide with small losses in *EDPW* and vice versa), the total VaR will not fall below the corresponding *CPBP* VaR because of the non-negativity of losses. Considering the 99.9% VaR of roughly \$3.6T for *CPBP* and around \$0.12T for *EDPW*, the scope of VaR reduction is small on a relative scale.

The frequency model indifference mentioned above, highlights the importance of an accurate severity model. As the 99.9% VaR is a consequence of a number of events in the severity distribution tail, the latter demands the practitioner's full attention in the modeling step. With regard to the surreal amounts of several trillion US Dollar, we would like to point out once more our scepticism in the correctness of the data material. Currently, there is no bank on earth which could handle extreme losses of that scale. Therefore, the results of this thesis have to be viewed from an abstract and more general perspective. If the outlier was discarded, the resulting minimum capital charge would have been much smaller than in the case considered in this thesis.[1] Assuming the accuracy of the data material – and no further information was submitted to us – this exclusion can not be justified.

[1]Due to the larger shape parameter, however, unrealistic results are also to be expected.

In fact, if the uncertainty about the outlier is ignored, the new developed scale-switching GPD model has proven to perform well in situations where an outlier deteriorates the overall tail fit. It managed to yield a sound tail behaviour and surpassed the usual GPD measured in terms of a likelihood-ratio test. Admittedly, the additional flexibility comes at the cost of increased model complexity and high estimation errors for the outlier scale parameter.

It should naturally be given some thought to alternative approaches. As the scale-switching GPD is clearly close to the very limits of what is statistically possible concerning the tail fit, one might even need to leave the terrain of statistics. A scenario analysis conducted by a committee of operational risk experts not only helps to rise the awareness of potential risks, but also might be more reliable than extrapolation methods.

Even a combination is possible: extreme scenarios are generated by the committee and used to overcome data scarcity in the tail distribution. Understandably it has to be taken into account that adding extreme events changes the overall proportions in the dataset. How this may look in practice, should be subject to further research. As the AMA comes with scenario analysis attached, not even an extra effort is necessary in that matter.

Chapter 7

Conclusions

The intention of this thesis has been to develop a practical method for dealing with operational loss data when a regular extreme value theory "Peaks over Threshold" approach fails to yield a reasonable fit. The decision of excluding the Advanced Measurement Approach from future considerations (probably starting from Basel IV), made by the Basel Committee on Banking Supervision, raised yet another more fundamental question. Is statistics able to model operational risk reliably?

Using monthly operational loss data from US businesses, two event types were looked upon, both having in common to feature heavy tails and one single extreme maximum loss. While a Generalised Pareto distribution could cope with the maximum in *EDPW*, it failed to fit the tail in *CPBP*. The scale of the maximum turned out to be in no relation to other tail events, which caused a poor modeling compromise. On the one hand regular tail events were underestimated since the outlier shifted the GPD towards more extreme events. On the other hand the uppermost tail was also underestimated, making the maximum loss highly unlikely to occur.

By relaxing the common assumption of independent and identically distributed random variables, it was shown how a tiny modification in the scale parameter can help ameliorating the overall model fit. The use of the additional scale parameter for huge outliers was justified by a significant increase in likelihood. However, one needs to bear in mind that the small sample size results in large estimation errors, particularly for the outlier scale parameter.

Having said this, it is in fact not surprising why the AMA was criticised to produce unreliable capital requirement estimates. The resulting Value-at-Risk would have been completely different if the outlier in *CPBP* had been dropped, if a standard GPD model

had been fitted or if, e.g., a Generalised Champernowne distribution had been fitted to the entire dataset. Even if data quality and data scarcity becomes less and less of a problem against the background of increasing awareness of operational risks, the nature of these risks makes modeling still a challenging problem.

Will the projected Standardised Measurement Approach overcome these issues? Actually, it solves the problem of allowing banks too much freedom in their model choice as well as it requires less effort regarding data collection and modeling. Furthermore, it makes it possible to draw comparisons between different financial institutions. On the other hand it might undermine the incentive of concentrating on understanding and examining operational risks.

Statistical research should not get discouraged by the decision of phasing out the AMA, but it should take it as a motivation to get an even better understanding of these risks and to choose more uncomfortable and unconventional paths in the future.

Appendices

Calculations

We want to give a formal proof for equation (4.2). Let X be a GPD random variable with parameters ξ and β. Since the distribution function and even the boundaries for x depend on the sign of ξ, we perform a case-by-case analysis.

Case $\xi > 0$:

$$\begin{aligned}\mathbb{E}[X] &= \int_0^\infty x dG_{\xi,\beta}(x) = \left[-x\left(1+\xi\frac{x}{\beta}\right)^{-1/\xi}\right]_0^\infty + \int_0^\infty \left(1+\xi\frac{x}{\beta}\right)^{-1/\xi} dx \\ &= \left[\frac{\beta}{\xi-1}\left(1+\xi\frac{x}{\beta}\right)^{1-1/\xi}\right]_0^\infty = \frac{\beta}{1-\xi}.\end{aligned} \tag{.1}$$

Case $\xi < 0$:

$$\begin{aligned}\mathbb{E}[X] &= \int_0^{-\beta/\xi} x dG_{\xi,\beta}(x) = \left[-x\left(1+\xi\frac{x}{\beta}\right)^{-1/\xi}\right]_0^{-\beta/\xi} + \int_0^{-\beta/\xi} \left(1+\xi\frac{x}{\beta}\right)^{-1/\xi} dx \\ &= \left[\frac{\beta}{\xi-1}\left(1+\xi\frac{x}{\beta}\right)^{1-1/\xi}\right]_0^{-\beta/\xi} = \frac{\beta}{1-\xi}.\end{aligned} \tag{.2}$$

Case $\xi = 0$:

$$\begin{aligned}\mathbb{E}[X] &= \int_0^\infty x dG_{\xi,\beta}(x) = \left[-x\exp\left(-\frac{x}{\beta}\right)\right]_0^\infty + \int_0^\infty \exp\left(-\frac{x}{\beta}\right) dx \\ &= \left[-\beta\exp\left(-\frac{x}{\beta}\right)\right]_0^\infty = \beta = \frac{\beta}{1-\xi}.\end{aligned} \tag{.3}$$

Bibliography

Theodore W Anderson and Donald A Darling. Asymptotic theory of certain" goodness of fit" criteria based on stochastic processes. *The annals of mathematical statistics*, pages 193–212, 1952.

Falko Aue and Michael Kalkbrenner. Lda at work: Deutsche bank's approach to quantifying operational risk. *Journal of Operational Risk*, 1(4):49–93, 2006.

Basel Committee on Banking Supervision. International convergence of capital measurement and capital standards. June 2006. URL `http://www.bis.org/publ/bcbs128.pdf`.

Basel Committee on Banking Supervision. Basel iii: A global regulatory framework for more resilient banks and banking systems. December 2010. URL `https://www.bis.org/publ/bcbs189_dec2010.pdf`. As at 19th April 2016.

Basel Committee on Banking Supervision. Principles for the sound management of operational risk. 2011. URL `http://www.bis.org/publ/bcbs195.pdf`.

Basel Committee on Banking Supervision. Fundamental review of the trading book: A revised market risk framework. October 2013. URL `http://www.bis.org/publ/bcbs265.pdf`. As at 22nd May 2016.

Basel Committee on Banking Supervision. A brief history of the basel committee. October 2015. URL `http://www.bis.org/bcbs/history.pdf`. As at 12th April 2016.

Basel Committee on Banking Supervision. Standardised measurement approach for operational risk. March 2016. URL `http://www.bis.org/bcbs/publ/d355.pdf`.

Patrick Billingsley. Probability and measure. wiley series in probability and mathematical statistics. 1995.

Catalina Bolancé, Montserrat Guillén, Jim Gustafsson, and Jens Perch Nielsen. *Quantitative Operational Risk Models*. CRC Press, 2012.

Bundesanstalt für Finanzdienstleistungsaufsicht. Banking supervision, 2016. URL `http://www.bafin.de/EN/DieBaFin/AufgabenGeschichte/Bankenaufsicht/bankingsupervision_node.html#doc2692256bodyText2`. As at 1st June 2016.

Bureau of Labor Statistics, United States Department of Labor. Historical Consumer Price Index for All Urban Consumers, Table 24, 2016. URL `http://www.bls.gov/cpi/cpid1512.pdf`. As at 9th February 2016.

Valérie Chavez-Demoulin, Paul Embrechts, and Johanna Nešlehová. Quantitative models for operational risk: extremes, dependence and aggregation. *Journal of Banking & Finance*, 30(10):2635–2658, 2006.

V Choulakian and MA Stephens. Goodness-of-fit tests for the generalized pareto distribution. *Technometrics*, 43(4):478–484, 2001.

Stuart Coles, Joanna Bawa, Lesley Trenner, and Pat Dorazio. *An introduction to statistical modeling of extreme values*, volume 208. Springer, 2001.

Marcelo Cruz. *Operational risk modelling and analysis: theory and practice*. Risk books, 2004.

Marcelo G Cruz. *Modeling, measuring and hedging operational risk*. Wiley Chichester, 2002.

Ralph B D'Agostino. *Goodness-of-fit-techniques*, volume 68. CRC press, 1986.

Stefano Demarta and Alexander J McNeil. The t copula and related copulas. *International Statistical Review/Revue Internationale de Statistique*, pages 111–129, 2005.

Annalisa Di Clemente and Claudio Romano. A copula-extreme value theory approach for modelling operational risk. *Operational Risk Modelling and Analysis*, pages 189–208, 2004.

Paul Embrechts and Giovanni Puccetti. Aggregating risk capital, with an application to operational risk. *The Geneva Risk and Insurance Review*, 31(2):71–90, 2006.

Paul Embrechts, Claudia Klüppelberg, and Thomas Mikosch. *Modelling extremal events*, volume 33. Springer Science & Business Media, 1997.

Ronald Aylmer Fisher and Leonard Henry Caleb Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24, pages 180–190. Cambridge Univ Press, 1928.

Antoine Frachot, Olivier Moudoulaud, Thierry Roncalli, et al. Loss distribution approach in practice, 2004.

Christian Genest, Bruno Rémillard, and David Beaudoin. Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and economics*, 44(2):199–213, 2009.

Boris Gnedenko. Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of mathematics*, pages 423–453, 1943.

Glyn A Holton. History of value-at-risk, 2002.

John C. Hull. *Risk Management and Financial Institutions*. John Wiley & Sons, 4 edition, 2015.

Andreas Jakobsson. *An Introduction to Time Series Modeling*. Studentlitteratur, 1 edition, 2013.

Arthur F Jenkinson. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81(348):158–171, 1955.

Harry Joe and James Jianmeng Xu. The estimation method of inference function for margins for multivariate models. Technical Report 166, Department of Statistics, University of British Columbia, 1996.

H Felix Kloman. Rethinking risk management. *Geneva Papers on Risk and Insurance. Issues and Practice*, pages 299–313, 1992.

Alexander J McNeil, Rüdiger Frey, and Paul Embrechts. *Quantitative risk management: Concepts, Techniques and Tools*. Princeton University Press, 2005.

Marco Moscadelli. The modelling of operational risk: experience with the analysis of the data collected by the basel committee. *Available at SSRN 557214*, 2004.

Roger B Nelsen. *An introduction to copulas*. Springer Science & Business Media, 2007.

Johanna Nešlehová, Paul Embrechts, and Valérie Chavez-Demoulin. Infinite mean models and the lda for operational risk. *Journal of Operational Risk*, 1(1):3–25, 2006.

Holger Rootzén and Nader Tajvidi. Extreme value statistics and wind storm losses: a case study. *Scandinavian Actuarial Journal*, 1997(1):70–94, 1997.

David Ruppert. *Statistics and data analysis for financial engineering*. Springer, 2011.

Michael A Stephens. Edf statistics for goodness of fit and some comparisons. *Journal of the American statistical Association*, 69(347):730–737, 1974.

Michael A Stephens. Goodness of fit for the extreme value distribution. *Biometrika*, 64 (3):583–588, 1977.

Richard Von Mises. La distribution de la plus grande de n valeurs. *Rev. math. Union interbalcanique*, 1(1), 1936.

Jun Yan et al. Enjoy the joy of copulas: with a package copula. *Journal of Statistical Software*, 21(4):1–21, 2007.