Operational Risk Modeling: Theory and Practice

J O H A N W A H L S T R Ö M

Master of Science Thesis
Stockholm, Sweden 2013

Operational Risk Modeling: Theory and Practice

J O H A N W A H L S T R Ö M

Master's Thesis in Mathematical Statistics (30 ECTS credits)
Master Programme in Applied and Computational Mathematics (120 credits)
Royal Institute of Technology year 2013
Supervisor at KTH was Filip Lindskog
Examiner was Filip Lindskog

TRITA-MAT-E 2013:58
ISRN-KTH/MAT/E--13/58-SE

Royal Institute of Technology
School of Engineering Sciences

KTH SCI
SE-100 44 Stockholm, Sweden

URL: www.kth.se/sci

Abstract

This thesis studies the Loss Distribution Approach for modeling of Operational Risk under Basel II from a practical and general perspective. Initial analysis supports the use of the Peaks over Threshold method for modeling the severity distributions of individual cells.

A method for weighting loss data subject to data capture bias is implemented and discussed. The idea of the method is that each loss event is registered if and only if it exceeds an outcome of a stochastic threshold. The method is shown to be very useful, but poses some challenges demanding the employment of qualitative reasoning.

The most well known estimators of both the extreme value threshold and the parameters in the Generalized Pareto Distribution are reviewed and studied from a theoretical perspective. We also introduce a GPD estimator which uses the Method-of-Moments estimate of the shape parameter while estimating the scale parameter by fitting a specific high quantile to empirical data. All estimators are then applied to available data sets and evaluated with respect to robustness and data fit.

We further review an analytical approximation of the regulatory capital for each cell and apply this to our model. The validity of the approximation is evaluated by using Monte Carlo estimates as a benchmark. This also leads us to study how the rate of convergence of the Monte Carlo estimates depends on the "heavy-tailedness" of the loss distribution.

A standard model for correlation between cells is discussed and explicit expressions limiting the actual correlation between the aggregated loss distributions in the model are presented. These bounds are then numerically estimated from data.

Acknowledgements

I would like to express my sincerest thanks to Mattias Larsson, Nordea, for constant encouragement and sharing of his industry expertise. My thanks also goes to Filip Lindskog, KTH, for valuable discussions and feedback regarding both technical aspects and the exposition of the thesis. I would further like to thank Ann-Charlotte Kjellberg and Eva Setzer-Fromell at SAS for providing me with data from the SAS OpRisk database. I am also grateful to Nina Sala, Nordea, for providing me with internal data and information regarding the collection process. Finally, I would like to thank Henrik Rahm, Nordea, for linking me to the project in the first place.

Stockholm, December 2013 Johan Wahlström

Contents

1 Introduction

1.1 Background

Even though operational risk is far from a new concept to anyone in the banking industry, it was for a long time seen as a risk that could be disregarded or neglected in comparison to credit- or market risk. The last decades of globalization and deregulation in the financial world has however brought about larger trading volumes and more diversity in the operating business of many companies and institutions, which has both increased the risks and the potential losses associated with operational risk. Examples include the "Big Bang" reform in Japan 1998, the Financial Services Act of 1999 and the expansion of the eurozone. This has also triggered the development of new complex financial products, designed to hedge newly emerged risks or exploit markets that until recently have been illiquid. Simultaneously, technological innovations has enabled the growth of new services and activities such as online banking and high frequency trading. The increased speed and complexity of banking services and transactions has constantly driven the development of operational risk management and associated regulations forward, even though there exists many examples of lessons that had to be learned the hard way.

The Basel Committee did in 2001 define operational risk as

"The risk of direct or indirect loss resulting from inadequate or failed internal processes, people and systems or from external events."

Typical examples include fraud committed by employees, external individuals or organizations, compensation to employees, companies etc. due to damage to people, physical assets or the environment, and business disruption or losses connected to technical failures, human resources, accidental errors, terrorism or natural disasters. It is often said that operational risk, as opposed to credit-, market- or insurance risk, can be characterized by not being subject to speculation or other profit generating investments (for instance, the sellers of credit default swaps exploit credit risk for their own benefit). Still, the division is not always that clear cut since many companies are insured against losses attributed to operational risk.

The operational losses from a company can roughly be categorized into two main groups. These are losses with high frequency and low severity, and losses with low frequency and high severity. Losses with low frequency and low severity can often be omitted due to their obvious insignificance, while losses with high frequency and high severity for natural reasons do not exist. Most of the time, focus will lie on the low frequency/high severity-losses since these are very unpredictable and derives from risks which cannot be fully insured against. In worst case, the magnitude of a loss of this kind can be so large that it causes the company serious financial problems, or even leads the company to bankruptcy. We will below study some historical examples of these low-frequency losses.

When Barings Bank (London) declared bankruptcy in 1995, it was one of the oldest of its kind, having been founded in 1762. Nick Leeson, employed in 1989, had been assigned to perform low risk trading which would exploit arbitrage opportunities caused by price differences on exchanges in Japan and Singapore. Due to holding two positions at the same time (with trading *and* accounting duties), Leeson was first able to cover up the fact that he partly had abandoned his primary assignments, and instead had started to speculate by holding positions for a much longer time than intended. Later, this also allowed him to hide his losses until they were up to £827 million, much more than Barings total capital at the time.

In April 2010 the oil rig Deepwater Horizon, leased by the oil and gas company BP and located about 60 km south-east of the coast of Louisiana, exploded and caused the death of 11 employees. The explosion further caused the rig to sink, and enormous amounts of oil immediately started to leak into the Gulf of

Mexico. The incident led to huge losses for the fishing and tourism industries, and is by many regarded as the worst man-made environmental disaster in the US to date. As of February 2013, BP had paid out $42 billion in compensations, but the total economic loss for BP is expected to follow from the conclusion of the still ongoing legal proceedings.

In the same month, a volcano eruption in Iceland caused an ash cloud to be formed and sent into the upper part of the atmosphere. Previous experiences had shown that volcanic ash had the potential to damage aircraft engines in air, and since no adequate tests of this effect had been performed, airspace regulators decided to cancel nearly all flights in northern Europe for more than a week. The biggest economic losses following the event could be found in the airline industry and industries largely dependent on importing/exporting, not to mention lost revenue as a consequence of the many delays and cancellations of cultural events and meetings.

1.2 Regulations

The Basel Committee of Banking Supervision was established in 1974, with the aim of stabilizing banking and currency markets. The Committee released the Capital Accord, now commonly referred to as *Basel I*, in 1988. The accord primarily regulated credit risk, although also other types of risks were implicitly covered. Market risk was explicitly included in the updated guidelines, released in 1996, and two years later, drafts of *Basel II* were published.

The first mention of any capital requirement directly related to operational risk was in January 2001 when the Basel Committee publicized a consultative document focusing on the subject. Basel II, a more complete, flexible and modern framework in comparison to its predecessor, was released in 2004, while minor changes and additions were published during the subsequent years.

The committee has no formal legal authority, and the aim of the guidelines is merely to formulate broad standards which encourage convergence of risk measurement approaches, while at the same time allowing for different, locally tailored implementations. It is then the responsibility of national central banks and other institutions to decide on how to carry out and regulate these frameworks, in whichever form they consider most suitable for their specific needs and circumstances. *Basel III* was agreed upon in 2011 and is expected to be fully implemented in 2018. The European implementation have been constructed by the European Commission, and is called the *Capital Requirements Directive IV*.

The Basel accord consists of three pillars. The first one deals with credit, market and operational risk, while the second describes how banks and regulators should assess the risks from the first pillar, and also acknowledges risks not covered by the first pillar. The last pillar aims to encourage transparency of the process in which corporations meet the requirements of the first and the second pillar.

The first pillar describes three approaches of varying complexity and sensitivity for the modeling of operational risk. The most primitive is the *Basic Indicator Approach* in which the regulatory capital is calculated as a percentage of the average gross income from the last three years (years with negative gross income excluded). The *Standardized Approach* is somewhat more detailed, and entails the calculation of gross income for several different business lines. These numbers are then multiplied by specific factors for each business line which give the respective capital charges, and the total capital requirement is simply the sum of the charges for the individual business lines.

This thesis will focus on the *Advanced Measurement Approach* (AMA), which allows banks to themselves model operational risks under loose guidelines provided by the Basel committee. AMA is ordinarily implemented using the *Loss Distribution Approach* (LDA) described in section 2.1. International banks with "significant risk exposures" are expected to implement AMA and cannot go back to a simpler approach once AMA has been implemented (see paragraphs 647 and 648 in International Convergence of

Capital Measurement and Capital Standards (2006)).

1.3 Statement of Purpose

Operational risk modeling involves some very specific challenges related to robustness. First of all, the available data is often insufficient and unreliable. While the Basel guidelines encourages an estimation of the 0.999-quantile of the yearly aggregated loss, no bank has access to anything close to a thousand observations of annual losses with relevant risk characteristics. Furthermore, losses from databases such as ORX seldom specify which bank that reported a specific loss, and it is therefore in practice difficult to motivate any rejection of extreme losses. On top of this, the accounting practices might differ between database members, which means that the BL-ET specification cannot be assumed to be identical amongst all members. The responsibility of mapping the losses into the correct business line and event type can seldom exclusively be handed to a specific expert team at each company, but will to some extent have to be performed by many different employees, which increases the risk of misspecifications.

Moreover, it is often neither reasonably nor desirable to increase the capital allocation related to specific business lines in an all too careless manner. For this reason, the thesis will particularly study how we can reduce the risk of any severe overestimation without losing sight of data sensitivity and modeling consistency. By constructing an efficient and robust simulation procedure, we hope to aid future estimations and inferences. Our aim can be summarized as providing an answer to the following:

How can modeling and implementation decisions help to improve the performance of the Loss Distribution Approach with respect to robustness and efficiency?

1.4 Outline of the Thesis

Chapter 1 serves to introduce the reader to the subject at hand and its historical as well as regulatory context. The object of the thesis is presented together with a brief summary of some of the main obstacles and challenges in operational risk modeling.

Chapter 2 reviews the necessary theory with an emphasis on practical problems of applying Extreme Value Theory (EVT) to operational risk models. The basis of the typical LDA model is described, together with the most commonly used probability distributions, analytical approximations of sought key figures, and some aspects of correlation modeling. Furthermore, some elementary robustness theory is studied and applied to Generalized Pareto Distribution (GPD) - estimators.

In Chapter 3, the available data is presented and we go through the process of filtering the data. A method to scale the probabilities of loss observations subject to data capture bias is reviewed and applied to our data sets.

Chapter 4 describes those analysis procedures employed in the subsequent chapter which requires a bit more detailed description. It introduces the MoMom-Q estimator and explains the implemented estimation and simulation procedures.

The main results of the thesis are presented in chapter 5. The chapter includes studies of both possible severity distributions, parameter estimators for the GPD, and different threshold estimators. One section studies the convergence rate of Monte Carlo (MC) estimates of capital requirements, and another compares the MC estimates to analytical approximations. The robustness with respect to outliers of high severity, as well as the stability in terms of the threshold choice, is studied for several different parameter estimators. The next to last section discusses cell aggregation from a practical and regulatory perspective. Finally, theoretical bounds on the correlation of aggregated loss distributions given a specific correlation

model in the LDA is applied to the final model of the thesis. This then gives some numerical results supporting the notion that the assumption of perfect correlation in between cells is all too conservative with the given correlation model.

The results are then summarized in Chapter 6 along with some general conclusions that can be drawn from the study. At last, Chapter 7 gives some suggestions on how further studies of the subject could proceed, while also pointing out some of the limitations of the thesis.

2 Preliminary Theory

2.1 The Loss Distribution Approach

Let us assume that we have divided our observations into business lines (BLs) and event types (ETs). The total loss in cell (i, j), i.e. the losses in business line i emerging from event type j during the time interval $[t, t + \tau]$, is determined by the independent stochastic variables $N_{i,j}$ (the total number of losses in the cell) and $X^k_{i,j}$ (the size of the k:th loss in the cell) and can be written as

$$S_{i,j} = \sum_{k=1}^{N_{i,j}} X^k_{i,j}.$$

From here on, the indexes denoting the cell in question will be omitted due to notational convenience.

We will then denote the probability distribution function (pdf) of the loss frequency of a given cell by $f_{ev}(n)$, where $n \in \mathbb{N} \cup \{0\}$. Similarly, the pdf of the loss severity in the same cell can be written as $f_{sev}(x)$, where obviously $f_{sev}(x) = 0$ for $x < 0$. Furthermore, we can express the pdf of S as

$$f_{agg}(s) = \begin{cases} \sum_{n=1}^{\infty} f_{ev}(n) f^{n*}_{sev}(s) & , s > 0 \\ f_{ev}(0) & , s = 0 \, , \end{cases}$$

where $f^{n*}_{sev}(x)$ is the n-fold convolution of $f_{sev}(x)$ with itself, i.e.

$$f^{n*}_{sev}(x) = \begin{cases} f^{1*}_{sev}(x) = f_{sev}(x) \\ f^{2*}_{sev}(x) = f_{sev}(x) * f_{sev}(x) = \int_{-\infty}^{\infty} f_{sev}(x - y) f_{sev}(y) dy \\ f^{n*}_{sev}(x) = f^{(n-1)*}_{sev}(x) * f_{sev}(x), \; n > 2 \end{cases}$$

We will use $\tau = 1$ year throughout the thesis.

2.2 Risk Measures

The guidelines from the Basel committee does *not* specify which risk measure to use in AMA. We will therefore present all three measures which are mentioned in the guidelines (see paragraph 220 in Operational Risk - Supervisory Guidelines for the Advanced Measurement Approaches).

The most well known risk measure today is Value-at-Risk (VaR). To illustrate the use of VaR, consider the stochastic variable L which represents the loss from some given investment during the time interval $[t, t + \tau]$. $\mathrm{VaR}_\alpha(L)$ then equals the smallest threshold value l for which the probability that L exceeds l is not greater than the confidence level α. Mathematically, this can be written as

$$\mathrm{VaR}_\alpha(L) = \inf \{l : \mathrm{P}(L > l) \leq \alpha\} = \inf \{l : F_L(l) \geq 1 - \alpha\},$$

which obviously is equal to $F^{-1}_L(1 - \alpha)$ when F_L is continuous and strictly increasing (as should be the case under most practical circumstances). You should be aware of the fact that other definitions of VaR might denote this by $\mathrm{VaR}_{1-\alpha}(L)$ and that there also exists definitions which discounts the prevailing risk-free interest rate during $[t, t + \tau]$ from the value of the VaR.

VaR is often criticized since it completely ignores the shape of the distribution beyond the chosen confidence level. With this in mind, it is not hard to imagine two loss distributions with identical VaR whose implied risks differ greatly. This motivates the introduction of Expected Shortfall (ES), sometimes also called Conditional Value at Risk (CVaR). $\mathrm{ES}_\alpha(L)$ is the average value of $\mathrm{VaR}_\beta(L)$ given $0 \leq \beta \leq \alpha$ (with all β:s given the same weight), i.e.

$$\mathrm{ES}_\alpha(L) = \frac{1}{\alpha} \int_0^\alpha \mathrm{VaR}_\beta(L) d\beta.$$

From the definition, it should be clear that it is not possible to "hide risk in the tail" using ES, as opposed to when using VaR. For this and other reasons, ES has lately been gaining more recognition at the expense of VaR.

Another practical risk measure is Median Shortfall (MS). Median shortfall is simply the capital that needs to put away to ensure that you with probability 1/2 will cover all losses above some threshold. This can be expressed as

$$\mathrm{MS}_\alpha(L) = \mathrm{VaR}_\alpha(L) + \inf\left\{l : \mathrm{P}(L - \mathrm{VaR}_\alpha(L) \leq l | L > \mathrm{VaR}_\alpha(L)) \geq 1/2\right\}.$$

2.3 The Poisson Distribution

The Poisson distribution is the standard choice of frequency distribution. This can be motivated by the fact that the length of time between two events should be exponentially distributed, since the exponential distribution is the unique distribution that is synonymous with "lack of memory" (see theorem 2.2 in Enger and Grandell (2006), i.e.

$$P(X > x + y \,|X > y) = P(X > x) \Longleftrightarrow X \in \mathrm{Exp}(\lambda).$$

In other words, the fact that no internal fraud was reported last month does not mean that the probability suddenly is greater (or smaller), ceteris paribus, that an internal fraud will be reported this month. This means that the number of reported events starting from some time t_0 can be described by a Poisson process, and further that the total number of losses in the interval $[t, t + \tau]$ is Poisson distributed (see section 8.1 in Gut (2009)).

2.4 Extreme Value Theory

EVT is often employed when estimating the severity distribution. This section gives a short summary of the theoretical foundation of the most commonly used techniques.

Definition: Consider the i.i.d. variables, $X, X_1, ..., X_n$, and define the variable M_n by $M_n = \max(X_1, ..., X_n)$. If there exists $c_n \in (0, \infty)$ and $d_n \in (-\infty, \infty)$ such that

$$\lim_{n\to\infty} \mathrm{P}\left(\frac{M_n - d_n}{c_n} \leq x\right) = F_H(x),$$

where $F_H(x)$ is the cumulative distribution function (cdf) of some variable H, then X is said to be in the *Maximum Domain of Attraction* of H (written $X \in MDA(H)$).

First Theorem in Extreme Value Theory (the Fisher-Tippett-Gnedenko Theorem): If $X \in MDA(H)$, then H belongs to the *Generalized Extreme Value* (GEV) distribution, defined by the cdf

$$F_{H_\xi}(x) = \begin{cases} \exp\left\{-(1+\xi x)^{-1/\xi}\right\} & , \xi \neq 0 \\ \exp\left\{-e^{-x}\right\} & , \xi = 0, \end{cases}$$

for $1 + \xi x > 0$ and $\xi \in \mathbb{R}$.

Remark 1: Notice that the support of H_ξ, i.e. $\left\{x \in \mathbb{R} : f_{H_\xi}(x) \neq 0\right\}$, is

$$\begin{cases} x > -\frac{1}{\xi} & , \xi > 0 \\ x < -\frac{1}{\xi} & , \xi < 0 \\ x \in \mathbb{R} & , \xi = 0. \end{cases}$$

Remark 2: It is reasonable to assume that all standard continuous distributions belong to $MDA(H_\xi)$ (See Embrechts et al. (2005), section 7.1.2).

Second Theorem in Extreme Value Theory (the Pickands-Balkema-de Haan Theorem): Assume that $X, X_1, ..., X_n$ are i.i.d. and belongs to the domain of attraction of some H, and denote the conditional excess distribution function by $F_u(x) = \mathrm{P}(X - u \leq x | X > u)$ for some $u \in \mathbb{R}$. It then holds that

$$\lim_{u \uparrow x_F} \sup_{0 < x < x_F - u} |F_u(x) - G_{\xi,\beta}(x)| = 0,$$

where $G_{\xi,\beta}$ denotes the *Generalized Pareto Distribution* (GPD) function, i.e.

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

for

$$\begin{cases} 0 \leq x & , \xi \geq 0 \\ 0 \leq x \leq -\frac{\beta}{\xi} & , \xi < 0, \end{cases}$$

$\beta \in (0, \infty)$ (the scale parameter) and $\xi \in (-\infty, \infty)$ (the shape parameter). x_F is the right endpoint of X, i.e. $x_F = \sup\{x \in \mathbb{R} : F(x) < 1\}$.

The *Block Maxima Method* is a straightforward application of the first fundamental theorem in EVT where one will fit the GEV distribution to data consisting of the largest losses from some specified equidistant time intervals. This method is seldom used when modeling operational risk, since it rejects almost all data points and take no consideration to the distribution of the aggregated losses. A more practical method is *Peaks Over Threshold* (POT), in which you choose some convenient threshold u and try to fit the GPD to all loss data which exceeds u. The typical choice for the estimation of the severity distribution below the threshold is the Empirical Distribution Function (edf), in general defined by

$$\hat{F}_X(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{x_i < x\},$$

where $x_1, ..., x_n$ are observations of the stochastic variable X and $\mathbf{1}(\cdot)$ denotes the indicator function.

2.5 Approximations of Risk Measures

Böcker and Klüpperberg (2005) noted that the VaR of an aggregated loss variable S with the associated severity cdf $F_{sev}(x)$ can be approximated by

$$\mathrm{VaR}_\alpha(S) \approx F_{sev}^{-1}\left(1 - \frac{\alpha}{\mathrm{E}[N]}\right), \tag{1}$$

given that F_{sev} is subexponential (see definition 1.3.3 in Embrechts et al. (1997)) and

$$\sum_{n=0}^{\infty}(1+\epsilon)^n f_{ev}(n) < \infty, \tag{2}$$

for some $\epsilon > 0$. With the modeling of $F_{sev}(x)$ as a piecewise distribution with an empirical body and a GPD in the upper tail, this can be estimated by (see section 8.2 in the appendix)

$$\mathrm{VaR}_\alpha(S) \approx u + \frac{\beta}{\xi}\left(\left(\frac{N_{losses>u}}{N_{losses}}\frac{\mathrm{E}[N]}{\alpha}\right)^{\xi} - 1\right), \tag{3}$$

where N_{losses} denotes the total number of observations and $N_{losses>u}$ the number of observations exceeding the threshold u. In the same manner the expected shortfall can be estimated by

$$\mathrm{ES}_\alpha(S) \approx u - \frac{\beta}{\xi} + \frac{\beta}{\xi(1-\xi)}\left(\frac{N_{losses>u}}{N_{losses}}\frac{\mathrm{E}[N]}{\alpha}\right)^{\xi}.$$

2.6 Parameter Estimation for the Generalized Pareto Distribution

This section will review some methods for estimating the parameters in the GPD. It is generally harder to estimate the shape parameter than the scale parameter, why research has focused on the former problem. Also notice that with ξ given, β will only *scale* the independent variable in the pdf. Throughout the section we will assume that we have access to n ordered observations $x_n \leq ... \leq x_2 \leq x_1$, which are exceedances (over some threshold u), derived from a larger set of observations.

2.6.1 Hill's Estimator

Hill (1975) introduced what would become known as the Hill estimator,

$$\hat{\xi}_{H_k} = \frac{1}{k}\sum_{i=1}^{k}\ln(x_i) - \ln(x_{k+1}),$$

$$\hat{\beta}_{H_k} = \frac{x_{k+1}}{\hat{\xi}_{H_k}}\left(\frac{k}{n}\right)^{\hat{\xi}_{H_k}}.$$

Beirlant et al. (2004) presents four natural ways to introduce this estimator, all based on the upper tail behaviour of the GPD. As should be intuitively clear, the bias in the estimate of ξ will increase with k while the variance of the estimate decreases (see page 341 in Embrechts et al. (1997)).

2.6.2 Pickands' Estimator

Pickands (1975) proposed the estimators

$$\hat{\xi}_{P_k} = \frac{1}{\ln 2} \ln \left(\frac{x_k - x_{2k}}{x_{2k} - x_{4k}} \right),$$

and

$$\hat{\beta}_{P_k} = \frac{x_{2k} - x_{4k}}{\ln 2 \int_0^{\hat{\xi}_{P_k} s} e^{\hat{\xi}_{P_k} s} ds},$$

for some $k \in \{1, ..., \lfloor n/4 \rfloor\}$. The estimators are obtained by "matching" theoretical and empirical quantiles. The primary downside with Pickands' Estimator is that most observations are discarded, and hence convergence will be slow.

Both $\hat{\xi}_{H_k}$ and $\hat{\xi}_{P_k}$ converge to ξ in probability when $k, n \to \infty$ and $k/n \to 0$ (see Theorems 6.4.1 and 6.4.6 in Embrechts et al. (1997)).

2.6.3 The Maximum Likelihood Estimator

With $\xi \neq 0$, we have $\frac{dG_{\xi,\beta}(x)}{dx} = g_{\xi,\beta}(x) = \frac{1}{\beta}(1 + \xi \frac{x}{\beta})^{-1/\xi - 1}$, which gives the log likelihood function

$$\ln l(\xi, \beta) = -n \ln(\beta) - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{n} \ln(1 + \xi \frac{x_i}{\beta}). \tag{4}$$

By introducing $\tau = -\xi/\beta$, and substituting β with $-\xi/\tau$ we get

$$\ln l(\xi, \tau) = -n \ln(-\frac{\xi}{\tau}) - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{n} \ln(1 - \tau x_i),$$

subject to $\tau < 1/x_1$ (which can be derived from the upper bound on the independent variable when $\xi < 0$) and $\xi \geq -1$ (the likelihood function is unbounded when $\xi < -1$). The requirement $\frac{\partial \ln l(\xi,\tau)}{\partial \xi} = 0$ now gives

$$\xi = \frac{1}{n} \sum_{i=1}^{n} \ln(1 - \tau x_i).$$

By maximizing $l(\xi(\tau), \tau)$ (which has to be done using numerical methods), we can then obtain $\hat{\tau} = \arg\max_{\tau} l(\xi(\tau), \tau)$, which further gives us the estimates

$$\hat{\xi}_{ML} = \frac{1}{n} \sum_{i=1}^{n} \ln(1 - \hat{\tau} x_i),$$

and

$$\hat{\beta}_{ML} = -\frac{\hat{\xi}_{ML}}{\hat{\tau}}.$$

There is a consensus that the maximum likelihood (ML) estimator is well performing in the presence of large samples. Unfortunately, the scarcity of data is often severe in practice, in which case other estimators have proven more effective (see Deidda and Puliga (2009)).

2.6.4 Huisman's Estimator

In Huisman et al. (2001) it is noted that the bias of the Hill estimator is approximately linear in k when k is sufficiently small. This motivates the introduction of the regression model

$$\hat{\xi}_{H_k} = \beta_0 + \beta_1 k + \epsilon_k, \quad k \in 1, ..., \kappa,$$

where β_0 is the sought after estimate of ξ, and ϵ_k are the error terms. Due to dependence between the estimates (the complete set of data points used for $\hat{\xi}_{H_k}$ will also be used for $\hat{\xi}_{H_{k+1}}$ etc.) the model is heteroscedastic, and the standard ordinary least squares estimate is therefore usually replaced by a weighted least squares estimate which gives the weight $\sqrt{k}$ to each equation (the standard deviation of $\hat{\xi}_{H_k}$ is inversely proportional to $\sqrt{k}$). This can be shown to yield the multilinear estimate

$$\hat{\xi}_{Hu_\kappa} = \hat{\beta}_0(\kappa) = \sum_{k=1}^{\kappa} w_k(\kappa)\hat{\xi}_{H_k},$$

where $w_k(\kappa)$ are some constants only dependent on k and κ.

2.6.5 The Method-of-Moments Estimator

Hosking and Wallis (1987) were the first to derive the Method-of-Moments (MoMom) estimator. Provided that $\xi < \frac{1}{2}$ and $\xi \neq 0$ (the case $\xi = 0$ is non-relevant in practice), the mean, μ, and the variance, σ^2, of the GPD can be written as (see sections 8.3 and 8.4 in the appendix)

$$\mu = \frac{\beta}{1-\xi},$$

and

$$\sigma^2 = \frac{\beta^2}{(1-\xi)^2(1-2\xi)}.$$

The estimates of the shape and the scale parameter can now be defined as the parameters which allows the theoretical mean and variance to equal the sample mean and variance respectively, i.e.

$$\hat{\xi}_{MoM} = \frac{1}{2}\left(1 - \frac{\hat{\mu}^2}{\hat{\sigma}^2}\right),$$
$$\hat{\beta}_{MoM} = \frac{1}{2}\hat{\mu}\left(1 + \frac{\hat{\mu}^2}{\hat{\sigma}^2}\right),$$

where

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

and

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu})^2.$$

2.6.6 The Method-of-Probability-Weighted-Moments Estimator

Also introduced in Hosking and Wallis (1987), the Method-of-Probability-Weighted-Moments (MoPW-Mom) estimator is based on the same principle as the MoMom estimator. The estimator attempts to fit the theoretical probability weighted moments, defined by

$$M_{p,r,s} = \mathrm{E}[X^p (F(X))^r (1 - F(X))^s],$$

for some random variable X with cdf $F(x)$, with the corresponding sample estimates. For the GPD, assuming $\xi < 1$, it is especially convenient and simple to use

$$M_{1,0,0} = \mathrm{E}[X] = \frac{\beta}{1-\xi},$$

and (see section 8.5 in the appendix)

$$M_{1,0,1} = \mathrm{E}[X(1 - F(X))] = \frac{\beta}{2(2-\xi)},$$

which analogously to the MoMom estimator gives

$$\hat{\xi}_{PWM} = 2 - \frac{\hat{M}_{1,0,0}}{\hat{M}_{1,0,0} - 2\hat{M}_{1,0,1}},$$

and

$$\hat{\beta}_{PWM} = \frac{2\hat{M}_{1,0,0}\hat{M}_{1,0,1}}{\hat{M}_{1,0,0} - 2\hat{M}_{1,0,1}},$$

with the unbiased empirical estimator

$$\hat{M}_{1,0,s} = \frac{(n-s-1)!}{n!}\sum_{i=1}^{n}(n-i)(n-i-1)...(n-i-s+1)x_{n+1-i}.$$

Notice that $\hat{\xi}_{PWM} < 1$ (excluding the unrealistic case that all losses are identical in value), since

$$\hat{\xi}_{PWM} = \frac{\hat{M}_{1,0,0} - 4\hat{M}_{1,0,1}}{\hat{M}_{1,0,0} - 2\hat{M}_{1,0,1}},$$

and

$$\hat{M}_{1,0,0} - 2\hat{M}_{1,0,1} = \frac{1}{n}\sum_{i=1}^{n}\left(1 - 2\frac{n-i}{n-1}\right)x_{n+1-i} = \frac{1}{n(n-1)}\sum_{i=1}^{n}(2i-(n+1))x_{n+1-i}$$
$$= \frac{1}{n(n-1)}((n-1)x_1 + (n-3)x_2 + ... - (n-3)x_{n-1} - (n-1)x_n) > 0,$$

whenever we don't have $x_1 = x_2 = ... = x_n$.

2.6.7 The Method-of-Medians Estimator

The Method of Medians Estimator was first applied to the GPD in Peng and Welsh (2001), and is defined as the solution to

$$\text{Median}\{x_i\} = G_{\hat{\xi},\hat{\beta}}(0.5)^{-1} = \frac{\hat{\beta}}{\hat{\xi}}(2^{\hat{\xi}} - 1), \tag{5}$$
$$\text{Median}\left\{\frac{\ln(1+\hat{\xi}x_i/\hat{\beta})}{\hat{\xi}^2} - \frac{(1+\hat{\xi})x_i}{\hat{\beta}\hat{\xi} + \hat{\xi}^2 x_i}\right\} = z(\hat{\xi}),$$

where $z(\hat{\xi})$ is defined by

$$\int_{\Omega} dy = 1/2, \quad \Omega = \{0 < y < 1, -\frac{\ln y}{\hat{\xi}} - \frac{1+\hat{\xi}}{\hat{\xi}^2}(1 - y^{\hat{\xi}}) > z(\hat{\xi})\}.$$

The estimator is obtained by equating the theoretical and empirical score function (i.e. the gradient of the likelihood function with respect to the parameters).

2.6.8 The kMedMad Estimator

The kMedMad estimator is a special kind of Location-Dispersion (LD) estimator. The LD estimators were introduced in Marazzi and Ruffieux (1998) and simply attempts to fit some chosen theoretical and observed measures of location and dispersion. The kMedMad specifically uses the median and the k-Median-of-Absolute-Deviations (kMad), the last of which is defined by,

$$\inf\{t > 0 : F_X(F_X^{-1}(0.5) + kt) - F(F_X^{-1}(0.5) - t) \geq 1/2\},$$

for some $k > 0$. kMedMad is a generalization of MedMad $= \text{kMedMad}\big|_{k=1}$, both of which was introduced in Ruckdeschel and Horbenko (2010), the latter with the object of improving the finite sample breakdown point (see section 2.9.5) of the former. It is straightforward to show that this gives the estimates as the solution to equation (5) and

$$\left(1 + \hat{\xi}\frac{\frac{\hat{\beta}}{\hat{\xi}}(2^{\hat{\xi}} - 1) - t^*}{\hat{\beta}}\right)^{-1/\hat{\xi}} - \left(1 + \hat{\xi}\frac{\frac{\hat{\beta}}{\hat{\xi}}(2^{\hat{\xi}} - 1) + kt^*}{\hat{\beta}}\right)^{-1/\hat{\xi}} = \frac{1}{2},$$

where

$$t^* = \inf\{t > 0 : \hat{F}(\text{median}\{x_i\} + kt) - \hat{F}((\text{median}\{x_i\}) - t) \geq 1/2\}.$$

Typically, we obtain $k > 1$ when k is optimized with respect some robustness criteria. This choice of k will counteract the natural tendency of the MedMad estimator to give more weight to smaller observations due to the asymmetry of the GPD.

2.7 Threshold Estimation

This section will review some common methods to estimate the threshold when using the POT method. The data $x_n \leq ... \leq x_2 \leq x_1$ should here be thought of as the complete original set of observations, rather than exceedances above some threshold as in the last section.

2.7.1 The Mean Excess Plot

The mean excess function is defined as $e(u) = E[X - u|X > u]$. For the GPD, it evaluates to (see section 8.6 in the appendix)

$$e(u) = \frac{\beta}{1-\xi} + u\frac{\xi}{1-\xi}, \tag{6}$$

for $\xi < 1$ and $\xi \neq 0$.

By plotting the empirical mean excess function, $\hat{e}(u)$, it is possible to graphically choose some threshold u' where $\hat{e}(u)$ is approximately linear for $u > u'$. The points which are typically plotted are $(x_i, \hat{e}(x_i))$, for $i = 2, 3, ...$, where

$$\hat{e}(u) = \frac{\sum_{i=1}^{n}(x_i - u)\,\mathbf{1}(x_i > u)}{\sum_{i=1}^{n}\mathbf{1}(x_i > u)}.$$

2.7.2 The Median Excess Plot

One might also characterize a probability distribution by its median excess function $f(u) = \left(F_X^{(u)}\right)^{-1/2}(1/2)$, where $F_X^{(u)}(x) = P(X - u \leq x|X > u)$. For the GPD we have (see section 8.7 in the appendix)

$$f(u) = \frac{\beta}{\xi}(2^{\xi} - 1) + u(2^{\xi} - 1). \tag{7}$$

Practitioners use the median excess plot to confirm the validity of a GPD-fit to data by plotting $(x_i, \hat{f}(x_i))$, for $i = 2, 3, ...$, where

$$\hat{f}(u) = \frac{x_{\lceil (k+1)/2 \rceil} + x_{\lfloor (k+1)/2 \rfloor}}{2} - u,$$

and $x_{k+1} \leq u < x_k$.

2.7.3 The Hill Plot

The Hill plot is simply the graph connecting the points $(k, \hat{\xi}_{H_k})$. Some studies subjectively chooses the threshold close to where the graph seems to become "unstable" (remember that the bias in the estimate will increase with k).

2.7.4 The Huisman Method

In Tursunalieva and Silvapulle (2011), the threshold is suggested to be chosen as the κ which minimizes $|\hat{\xi}_{H_\kappa} - \hat{\beta}_0(\kappa)|$. The Huisman method can be seen as giving an empirical estimate of the level where the total error from the bias in $\hat{\xi}_{H_\kappa}$ present for large values of k, and the large variance obtained with small values of k, is minimized. Since the idea of the method is to use the inherent k-dependence in the variance and bias of the Hill estimator, the Huisman Method is actually a formal implementation of the Hill plot.

2.7.5 The Riess-Thomas Method

Reiss and Thomas (2007), proposes that you select the number of extreme values as the k which minimizes

$$\mathrm{RT}_{k,\gamma} = \frac{1}{k}\sum_{i=1}^{k} i^{\gamma}|\hat{\xi}_i - \mathrm{median}\{\hat{\xi}_1, ..., \hat{\xi}_k\}|,$$

with $\gamma \in [0, 0.5]$, and where $\hat{\xi}_i$ denotes an estimate of the shape parameter obtained by using some chosen estimator and $x_{i+1} \leq u < x_i$. This is obviously a formal way of choosing the threshold at a level where the estimates are stable with respect to the threshold. They further suggest that you also try to minimize the alternative measure obtained by replacing the median in the sum by $\hat{\xi}_k$.

Some empirically motivated rules of thumb, which only uses the number of observations available have also been suggested. A number of these can be found in Scarrott and MacDonald (2012).

2.8 Severity Distributions

Aside from the GPD, we will also use two other commonly employed severity distributions: the *LogNormal Distribution* (LND) and the *Weibull Distribution* (WBD). A random variable X is said to be log-normally distributed if $X = e^Y$, where $Y \in N(\mu, \sigma)$. The ML estimates of the parameters μ and σ can be derived as (see section 8.8 in the appendix)

$$\hat{\mu} = \frac{\sum_{i=1}^{n} \ln x_i}{n}, \tag{8}$$

and

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (\ln x_i - \hat{\mu})^2}{n}. \tag{9}$$

Similarly, the WBD, defined by $F(x) = 1 - e^{\frac{x^k}{\lambda}}$ for $x \geq 0$, has the ML estimates $(\hat{\lambda}, \hat{k})$ given by (see section 8.9 in the appendix)

$$\frac{\sum_{i=1}^{n} x_i^{\hat{k}} \ln x_i}{\sum_{i=1}^{n} x_i^{\hat{k}}} - \frac{1}{\hat{k}} - \frac{1}{n}\sum_{i=1}^{n} \ln x_i = 0, \tag{10}$$

and

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} x_i^{\hat{k}}}{n}. \tag{11}$$

Since these distributions are heavy-tailed, i.e. have a cdf which approaches one slower than exponentially, they are popular choices for modeling loss data (which also often displays this characteristic).

2.9 Measures of Robustness

Robustness can intuitively be thought of as an estimators ability to limit the influence of outliers and data modifications. An estimator which amplifies negligible changes in in-data to extreme changes in out-data, can obviously not be called neither reliable nor robust, and its ability of making predictions should be doubted. This section will review some common measures of robustness suitable for analyzing previously mentioned estimators. First, we will study *local robustness*, i.e. how estimators withstand small deviations in the data. This analysis is traditionally carried out by using influence functions and other similar measures. We then examine the theory of *global robustness*, i.e. how estimators behave in the presence of unbounded outliers.

2.9.1 The Influence Function

Let $T(F)$ denote the limit in probability of the estimators $\{T_n\}$ given the cdf F. Most often, the estimators $\{T_n\}$ can be thought of as the values given by some estimator applied to the data points $\{x_1, ..., x_n\}$, which are outcomes from some stochastic variable. The *Influence Function* (IF) of T at x with respect to F, can then be defined as

$$\mathrm{IF}(x, T; F) = \lim_{\epsilon \downarrow 0} \frac{T((1-\epsilon)F + \epsilon H(x)) - T(F)}{\epsilon},$$

where $H(x)$ denotes the trivial distribution giving mass 1 to x. The IF should be thought of as the derivative of the influence on T of small impurities in observational data.

The IF of many of the GPD estimators introduced above can be found in Ruckdeschel and Horbenko (2010).

2.9.2 The Empirical Influence Function

If you want to study the IF of some estimator with regard to the cdf of some real life phenomena, you easily run into problems. The associated stochastic variable can probably not be described by any mathematical distribution function, and even if it could, the distribution is typically unknown to the observer. One alternative is then to use the *Empirical Influence Function* (EIF),

$$\mathrm{EIF}(x, T; \hat{F}_n) = \mathrm{IF}(x, T; \hat{F}_n),$$

where $\hat{F}_n$ denotes the edf of a set of observations $\{x_1, ..., x_n\}$. The EIF can under most circumstances be considered to be a reliable approximation of the influence function already for relatively small samples (see Opdyke and Cavallo (2012)). Theorems regarding requirements for asymptotic convergence of the empirical influence function to the corresponding influence function are beyond the scope of this thesis and the reader is referred to Nasser and Alam (2006) for technical details.

2.9.3 The Sensitivity Function

In many cases, due to scarcity of data, the asymptotic properties of estimators are not relevant, and it is more convenient to study their behaviour in the presence of finite samples when the proportional contamination of the data cannot be assumed to approach zero. This lead us to study the *Sensitivity Function* (SF),

$$\mathrm{SF}(x, X, \{T_n\}) = (n+1)(T_{n+1}(x_1, ..., x_n, x) - T_n(x_1, ..., x_n)),$$

where we use the notation $X = \{x_1, ..., x_n\}$, and where $T_n(x_1, ..., x_n)$ denotes the estimate from the estimator T_n given the data points $x_1, ..., x_n$. When $\mathrm{SF}(x, X, \{T_n\})$ is seen as a random variable dependent on the distribution F, most estimators satisfies

$$\lim_{n\to\infty} \mathrm{SF}(x, X, \{T_n\}) = \mathrm{IF}(x, T; F), \quad \forall\, x \in \mathbb{R}.$$

In Croux (1998), it is however showed that equality does not hold for estimators using the median of the observations.

2.9.4 The Breakdown Point

The *Breakdown Point*, $\epsilon^*(T; F)$, of an estimator T with respect to the cdf F is defined by

$$\epsilon^*(T; F) = \sup\{\epsilon \leq 1 : \sup_{d(F,F')<\epsilon} |T(F) - T(F')| < \infty\},$$

where we have used the Prohorov distance

$$d(F, F') = \inf\{\epsilon : F(x) \leq G(x_\epsilon) + \epsilon, \ \ \forall\, x, x_\epsilon \in \mathbb{R}, \ \ \text{s.t. } |x - x_\epsilon| < \epsilon\}.$$

If you think the definition is somewhat unintuitive, you can think of the breakdown point as the "distance" from F to the distribution closest to F for which the asymptotic value of the estimator T becomes unbounded.

Instead of studying how "close a parameter estimator is to becoming unbounded", it is in some cases more interesting to study how close it is to some given value which has a specific effect on the modeling. For instance, when fitting a GPD, it is interesting to see how much you need to change a given set of observations to get $\xi \geq 1/2$ (the variance becomes infinite) or $\xi \geq 1$ (the mean becomes infinite).

2.9.5 The Finite Sample Breakdown Point

The finite sample version of the breakdown point is simply called the *Finite Sample Breakdown Point* (FSBP), and is given by

$$\epsilon_n^*(X, T_n) = \frac{1}{n} \max\{m : \max_{i_1,...,i_m} \sup_{y_1,...,y_m} |T_n(z_n, ..., z_n)| < \infty\},$$

where $\{z_1, ..., z_n\}$ is obtained by replacing the data points $x_{i_1}, ..., x_{i_m}$ by $y_1, ..., y_m \in \mathbb{R}$ in $\{x_1, ..., x_n\}$. The FSBP can be said to be the largest fraction of the data $\{x_1, ..., x_n\}$ that one can change arbitrarily without risking unbounded estimates from T_n.

In general, those estimators which take many or all available observations into consideration will have a very small FSBP. For instance, all GPD estimators presented above which are not based on matching quantiles have a FSBP of $1/n$. Notice, however, that since the ξ estimates given by MoMom and MoPWMom are limited from above, their FSBP is 1 with respect to a positive infinite estimate of the shape

parameter. As for the other estimators, the FSBP of Pickands estimator and the kMedMad estimator can be found in Ruckdeschel and Horbenko (2010), while there is no known analytical expression for the FSBP of the MoMed estimator.

2.9.6 The Expected Finite Sample Breakdown Point

Motivated by the fact that $\epsilon_n^*(X, T_n)$ depends on $X = \{x_1, ..., x_n\}$ for some T_n, which severely restricts the possibilities of drawing any useful, general conclusion from the measure, Ruckdeschel and Horbenko (2012) introduced the *Expected Finite Sample Breakdown Point* (EFSBP), defined by

$$\bar{\epsilon}_n^*(T_n) = \mathrm{E}[\epsilon_n^*(X, T_n)].$$

Since $x_1, ..., x_n$ often are seen as outcomes of some cdf with the unknown parameter that T_n tries to estimate, the expectation is suggested to be taken as the expectation giving the smallest $\bar{\epsilon}_n^*$ under the assumption that the parameter belongs to some reasonable interval. In practice, one could for instance take the expectation under the assumption that the parameter estimate given by $T_n(x_1, ..., x_n)$ is equal to the true underlying parameter value.

2.10 Q-Q plots

A Q-Q plot displays the curve of a function $Q : [0, 1] \to \mathbb{R}^2$ defined by $Q(f) = (F_{X_1}^{-1}(f), F_{X_2}^{-1}(f))$, for some stochastic variables X_1 and X_2. The cdfs can either be edfs, cdfs of a distribution fitted to data or just some standard cdfs. Often times, one will fit a distribution to data and let the fitted distribution and the edf be drawn on the x- and y-axis respectively.

Q-Q plots are especially convenient when you want to examine and compare the tails of distributions. If the right part of the plot is convex, this means that the distribution on the y-axis has a heavier right tail than the distribution on the x-axis and vice versa.

2.11 Goodness-of-fit Tests

2.11.1 The Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov statistic associated with a cdf $F(x)$ and some edf $\hat{F}(x)$ is defined as

$$\mathrm{KS} = \sup_x |F(x) - \hat{F}(x)|. \tag{12}$$

Similarly, the two-sample Kolmogorov-Smirnov statistic of the edfs $\hat{F}_1(x)$ and $\hat{F}_2(x)$ is

$$\mathrm{KS}' = \sup_x |\hat{F}_1(x) - \hat{F}_2(x)|. \tag{13}$$

The associated p-values can then easily be calculated under the respective hypotheses that $F(x)$ is the true distribution of the sample associated with $\hat{F}(x)$, and that the samples resulting in $\hat{F}_1(x)$ and $\hat{F}_2(x)$ come from the same distribution.

2.11.2 The Upper Tail Anderson-Darling Statistic

The general quadratic edf statistic is defined by

$$\text{QEDF} = n \int_{-\infty}^{\infty} (\hat{F}(x) - F(x))^2 w(F(x)) dF(x),$$

and is a statistical tool to determine how well a distribution F fits the edf $\hat{F}$ given by a sample $x_n \leq ... \leq x_2 \leq x_1$. The weights $w(F(x)) = 1$ gives the Cramér-von Mises statistic, while $w(F(x)) = [F(x)(1-F(x))]^{-1}$ gives the Anderson-Darling statistic, which obviously emphasizes the lower and upper tail of the distribution. Chernobai et al. (2005) was the first paper to suggest the upper tail Anderson-Darling statistic (specifically introduced with applications related to operational risk in mind), which uses $w(F(x)) = (1 - F(x))^{-2}$. The statistic can easily be expressed in terms of the data points as (see section 8.10 in the appendix)

$$\text{UTAD} = 2 \sum_{i=1}^{n} \ln(1 - F(x_{n+1-i})) + \frac{1}{n} \sum_{i=1}^{n} (1 + 2(n-i)) \frac{1}{1 - F(x_{n+1-i})}. \tag{14}$$

Since the upper tail of the severity distribution more or less exclusively determines each cell's contribution to the total capital allocation, this is a very convenient statistic when analyzing suggested distributions.

2.12 Copula Theory

Copulas are used to model nonlinear dependence between stochastic variables. Let us consider a case were we are studying m variables, $X_1, ..., X_m$, with joint cdf $F(x_1, ..., x_m)$ and their respective one dimensional cdfs $F_1(x_1) = F(x_1, \infty, ..., \infty), ..., F_m(x_m) = F(\infty, ..., \infty, x_m)$. The associated copula $C : \mathbb{R}^m \mapsto [0, 1]$ is then defined by

$$F(x_1, ..., x_m) = C(F_1(x_1), ..., F_m(x_m)).$$

Sklar's Theorem says that the copula C always exists, and is unique whenever $F_1(x_1), ..., F_m(x_m)$ are continuous. Notice that the joint pdf can be expressed as

$$f(x_1, ..., x_m) = \frac{\partial^m F(x_1, ..., x_m)}{\partial x_1 \cdot ... \cdot \partial x_m} = f_1(x_1) \cdot ... \cdot f_m(x_m) \frac{\partial^m C(F_1(x_1), ..., F_m((x_m))}{\partial F_1 \cdot ... \cdot \partial F_m},$$

where $f_1(x_1), ..., f_m(x_m)$ of course denotes the respective pdfs of the stochastic variables.

Given a multidimensional sample $\{x_1^t, ..., x_m^t\}_{t=1}^T$, the *Empirical Copula* is defined by

$$\hat{C}\left(\frac{t_1}{T}, ..., \frac{t_m}{T}\right) = \frac{1}{T} \sum_{t=1}^{T} \mathbf{1}\{x_1^t \leq x_1^{(t_1)}, ..., x_m^t \leq x_m^{(t_m)}\},$$

for $1 \leq t_1, ..., t_m \leq T$, and where $x_j^{(1)} \leq x_j^{(2)} \leq .... \leq x_j^{(T)}, \ \ \forall\, j \in \{1, ..., m\}$.

The most well-known Copula is the Gaussian Copula. It is defined by

$$C_G(F_1, ..., F_m; \rho) = \Phi_\rho(\Phi^{-1}(F_1), ..., \Phi^{-1}(F_m)),$$

where Φ is the cdf of the standard one-dimensional normal distribution and Φ_ρ is the joint cdf of a multivariate normal distribution with the mean vector equal to the zero vector, and the correlation

matrix ρ. The correlation matrix ρ can be estimated using standard ML methods on the joint distribution function implied by the Copula (typically under some simplifying assumptions) or by minimizing some defined "distance" between the C_G and $\hat{C}$.

2.13 Correlation Modeling in the Loss Distribution Approach

There is no consensus on how to model and estimate correlation in the BL-ET matrix in LDA-models. Correlation models can be applied to the number of events in different cells, the severity distributions, or simply to the total yearly aggregated loss distributions. Other options include modeling of dependence "inside the cells" themselves, or common shock models where only the size of specific losses attributed to different cells are dependent. In the latter model, the idea is to capture the effects of singular events causing multiple losses of varying types, so called "split-losses". A review of available models together with further references can be found in Aue and Kalkbrenner (2006), chapter 7.

While most banks have to resort to models assuming perfect or near perfect correlation due to regulatory aspects and data insufficiency, Frachot et al. (2004) claims that this is in most cases leads to unrealistically conservative estimates. Assuming perfect correlation in between cells will furthermore punish corporations that use granular LDA models, which counteracts the idea that practitioners should be encouraged to use realistic models with great sensitivity to diverse risks in the operational business.

Let us consider two cells whose associated variables will be denoted with the indices 1 and 2 respectively. Further, assume that the number of events in the two cells are correlated and Poisson distributed with intensities λ_1 and λ_2 respectively, while the individual severities are independent on all levels. The correlation between the aggregated losses is dependent on the "event-correlation" through (see section 8.11 in the appendix)

$$\mathrm{Corr}(S_1, S_2) = \eta(X_1)\,\eta(X_2)\,\mathrm{Corr}(N_1, N_2), \tag{15}$$

where $\eta(X_i) = \mathrm{E}[X_i]/\sqrt{\mathrm{E}\left[X_i^2\right]}$ for $i = 1, 2$. By assuming a lognormal distribution for the severities and using data from Credit Lyonnais to estimate the distribution parameters, Frachot et al. (2004) uses (15) to argue that $\mathrm{Corr}(S_1, S_2) < 0.04$ in all practical cases. Similarly, we will in section 5.9 use parameter estimates to argue that the actual correlation between the yearly loss distributions of those cells that contribute the most to the total capital requirement in our model, assuming independence of individual severities, is far less than 1.

3 Data Material

The study have used data from three sources: the internal database, consisting of loss events from 2007 and onwards, the ORX database, i.e. data collected starting from 2002, gathered from banks with a membership in the international Operational Risk eXchange Association, and also, data collected by SAS Institute from publicly reported losses. All data have been categorized as belonging to one of ten BLs and one of seven ETs. The BLs are Corporate Finance, Trading & Sales, Retail Banking, Commercial Banking, Clearing, Agency Services, Asset Management, Retail Brokerage, Private Banking and Corporate Items, while the event types are Internal Fraud, External Fraud, Employment Practices, Clients & Products, Disasters & Public Safety, Technology & Infrastructure, and Execution & Delivery. There does exist a subdivision of BLs and ETs, however, it is seldom used in practice and will not be accounted for in this study. The BLs Private Banking and Corporate Items are mapped to Retail Banking and Retail Brokerage, respectively, in the Standardized Approach defined by Basel, and have merely been invented by ORX. A detailed account of the division into categories and other reporting issues can be found in Operational Risk Reporting Standards (2011).

3.1 Data Filtering

The ORX data was filtered to only include losses from Europe, since losses from other parts of the world are regarded as being subject to cultural and local effects, not relevant to our estimates. Examples include hierarchical organizational structures at banks affecting the amount of losses being "hidden" by employees, and differing legal frameworks across regions (for further examples see Graham (2008)). This is also supported by paragraph 250 in Operational Risk - Supervisory Guidelines for the Advanced Measurement Approaches (2011).

The SAS data was first filtered to only include losses from Europe categorized as belonging to Financial Services in the North American Industry Classification System. The losses were then converted from American dollars to euro (which is the currency of the ORX losses) using historical exchange rates from the year when the loss was recorded. Next, all SAS losses were matched against the ORX losses, after which duplicates were removed from the SAS data. A pair of duplicates were defined as losses found in the same cell in both sets, recorded in the same year, and where the loss difference was less than €100. Since the SAS losses only includes losses exceeding $100,000, this means that the relative loss difference in the two sets for the most part will have to be smaller than 1/1000 if the loss is to be removed. This is far from an infallible method due to the uncertainty of the correct exchange rate (the timestamp of the SAS losses only specifies the year), however, a more extensive removal of suspected duplicates would most certainly lead to incorrect removals.

All ORX data was attached with three dates: date of occurrence, date of discovery and date of recognition, where date of discovery is defined as "the date on which the firm became aware of the event" and date of recognition as "the date when a loss or reserve/provision was first recognized in the 'profit and loss'". In accordance with paragraph 29 in Operational Risk - Supervisory Guidelines for the Advanced Measurement Approaches (2011), the losses were classified according to discovery date. All losses were then adjusted with respect to the historical inflation of the euro area.

You should note that we have been forced to neglect several other biases related to time. Both the internal event data and the ORX losses will be affected by incomplete reporting practices during the initial years of reporting. This should generally result in fewer events and higher losses (with the assumption that lower losses are more easily neglected) during this period. Furthermore, changes in risk exposure over time can be handled by the use of *Business Environment and Internal Control Factors* (BEICFs,

see paragraphs 255 and 256 in Operational Risk - Supervisory Guidelines for the Advanced Measurement Approaches). They will, however, not be considered in this thesis. The delay in reporting might also have a lowering effect on the number of events reported for the last year. Nonetheless, the internal data does not indicate any severe bias due to this.

Every loss event from ORX specifies gross loss, net loss after direct recovery (which includes, for instance, amounts received in a legal settlement offsetting the initial loss) and net after all recovery (which also includes recoveries due to insurance payouts). Our model used the net loss after direct recovery, which, according to paragraph 95 in Operational Risk - Supervisory Guidelines for the Advanced Measurement Approaches (2011), is the most common approach when implementing AMA.

All losses except those from the BLs 0101, 0201, 0202, 0203, 0204, 0301, 0302, 0401, 0501, 0601, 0703, 0801 and 0901 have been discarded (using the standard notation in Operational Risk Reporting Standards (2011)). This is done since only these BLs are relevant for the internal business operations.

3.2 Data Weighting

Besides the truncations at €20,000 and $100,000 for the ORX and SAS data respectively, there are several other sources of bias in the external data. These include: *Scale Bias*, *Representation Bias* and *Data Capture Bias*.

Scale bias refers to the case of losses being proportional to the size of the associated organization or business area. While several propositions have been made about how to account for this bias, they were all disregarded in this thesis. First of all, regressions of the SAS losses on quantities relating to firm size did not reveal any significant relation between the variables (a result which agrees with the findings in Aue and Kalkbrenner (2006)), and further, no measures relating to firm size can be found in the ORX data, so the larger part of the losses would in either case have to remain unscaled.

The fact that neither members of ORX, nor the institutions labeled under "Financial Services" in the SAS data, are exposed to the exact same risks as those that we want to assess, is what gives rise to representation bias. Since the ORX members are only a limited number of major banks, while the filtered SAS losses are derived from a very diverse set of companies including pension funds, legal services and insurance carriers, the bias is most probably more severe for the SAS losses. Unfortunately, we cannot explicitly address this bias since we do not have access to any detailed information about the nature of the organization at which each specific loss occurred.

Data capture bias means that the probability of a loss being registered correctly will depend on the type and size of the loss. For instance, losses arising from a specific BL might have a larger tendency to be devaluated or swept under the carpet due to corporate policies or organizational culture. This is a problem which is hard to detect since it most probably exists in more or less all data, and hence do not bring about any natural benchmark.

The most critical bias that we face is believed to be the data capture bias arising from the registration procedure of the SAS data. The losses are drawn from a vast collection of publications, and while it is difficult to overview this process, it seems reasonably to suspect that the chance of a loss being publicly recognized increases with the size of the loss. This is further supported by the Q-Q plot of the ORX data versus the SAS data (see Figure 1 (left)). We handled this by undertaking the method outlined in Aue and Kalkbrenner (2006). The method rests on the assumption that the ORX data correctly reflects the internal losses and risks, something which was qualitatively motivated above, and also is quantitatively supported by the Q-Q plot in Figure 1 (right). Furthermore, we will model the severities of the ORX and the SAS losses with the stochastics variables X_{ORX} and X_{SAS}, related by $X_{SAS} = X_{ORX}|X_{ORX} \geq H$, where H is another stochastic variable representing the implicit threshold originating from the data

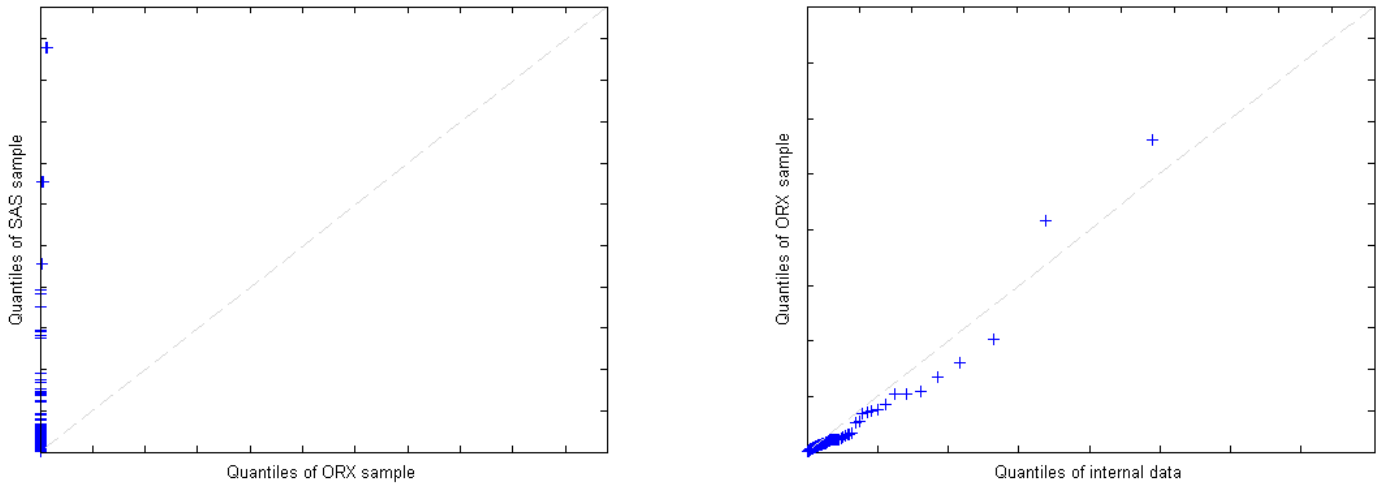


Figure 1 : *Q-Q plots based on the SAS and ORX samples (left), and the ORX and internal data samples (right). Only ORX and SAS losses exceeding €*120,000 *have been used in the left plot, and only internal losses exceeding €*20,000 *have been used in the right plot. The corresponding tick marks for the two axes represent the same values.*

capture bias. H will be presumed to follow the cdf $F_{H;\theta}(h)$ where θ is some set of parameters. The objective is then to minimize

$$\mathrm{Err}(\theta) = \sum_{i=1}^{k} (\mathrm{P}(H \leq X_{ORX} \leq S_i | H \leq X_{ORX}) - \mathrm{P}(X_{SAS} \leq S_i))^2,$$

with respect to θ, where $\{S_1, ..., S_k\}$ is some set of suitably chosen severities. Denoting the samples by $\{X^1_{ORX}, ..., X^{n_{ORX}}_{ORX}\}$ and $\{X^1_{SAS}, ..., X^{n_{SAS}}_{SAS}\}$, we used the estimate

$$\mathrm{P}(H \leq X_{ORX} \leq S_i | H \leq X_{ORX}) = \frac{\mathrm{P}(H \leq X_{ORX} \leq S_i \cap H \leq X_{ORX})}{\mathrm{P}(H \leq X_{ORX})}$$

$$= \frac{\mathrm{P}(X_{ORX} \leq S_i \cap H \leq X_{ORX})}{\mathrm{P}(H \leq X_{ORX})} \approx \frac{\sum_{X^j_{ORX} \leq S_i} F_{H;\theta}(X^j_{ORX})}{\sum_{j=1}^{n_{ORX}} F_{H;\theta}(X^j_{ORX})},$$

while $\mathrm{P}(X_{SAS} \leq S_i)$ was estimated using the edf. The samples consisted of all ORX and SAS losses exceeding €120,000. This threshold is introduced to diminish the time dependent effect of the actual SAS-threshold being larger than $100,000 due to inflation and varying historical exchange rates. The severities were chosen as

$$S_i = €120{,}000 \left(\frac{X^{max}_{ORX}}{€120{,}000} \right)^{(i-1)/(k-1)},$$

where X^{max}_{ORX} denotes the largest ORX loss and $k = 100$.

When θ has been estimated, a weighting of the probabilities of the SAS losses can be motivated by the following heuristic calculations:

$$\mathrm{P}(x < X_{SAS} < x + \Delta x) = \mathrm{P}(x < X_{ORX} < x + \Delta x | H < X_{ORX}) = \frac{\mathrm{P}(x < X_{ORX} < x + \Delta x \cap H < X_{ORX})}{\mathrm{P}(H < X_{ORX})}$$
$$= \frac{\mathrm{P}(x < X_{ORX} < x + \Delta x \cap H < x)}{\mathrm{P}(H < X_{ORX})} = \frac{\mathrm{P}(x < X_{ORX} < x + \Delta x)\mathrm{P}(H < x)}{\mathrm{P}(H < X_{ORX})}.$$

Since the denominator is independent of x, we can simply divide the probability of a SAS loss, X^i_{SAS}, in the edf by $F_{H;\theta}(X^i_{SAS})$ to adjust the probability for the impact of dismissing all losses falling below H (the probabilities will of course also have to be normalized). This will give a weighted edf,

$$\hat{F}_{SAS}(x) = \frac{\frac{1}{n}\sum_{X^i_{SAS}<x} 1/\mathrm{P}(H < X^i_{SAS})}{\frac{1}{n}\sum_{i=1}^{n_{SAS}} 1/\mathrm{P}(H < X^i_{SAS})}.$$

Notice that the weighting will not account for the original truncation at \$100,000.

As proposed in Aue and Kalkbrenner (2006), we modeled H using a log logistic distribution, i.e.

$$F_{H;\theta}(h) = \frac{1}{1 + (e^{\mu_H}/h)^{1/\sigma_H}},$$

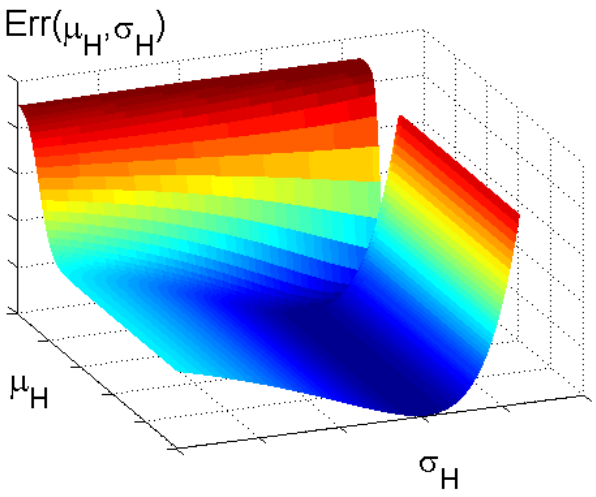


Figure 2 : *The characteristic dependence of the error function on the parameters μ_H and σ_H.*

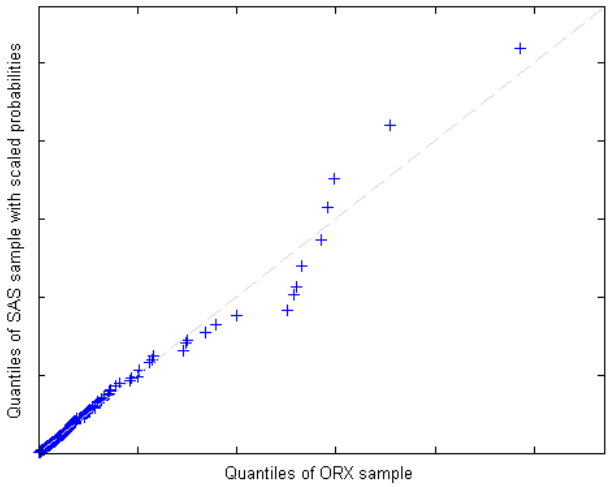


Figure 3 : *Q-Q plot based on the ORX and weighted SAS losses exceeding* €120,000.

which gave an excellent fit. Our observation is that this does *not* give any global minimum of $\mathrm{Err}(\mu_H, \sigma_H)$ when using our data sets. Instead, calculations indicate that the minimia exists at the end of a ribbon in the μ_H-σ_H-plane where $\mu_H \to \infty$. While the function value is very sensitive to changes in σ_H, it stabilizes quickly for large enough μ_H (see Figure 2). Changes in μ_H can however still be crucial for the scaling of the important high quantiles, even though the change in $\mathrm{Err}(\mu_H, \sigma_H)$ is negligible. This is due to the fact that

$$F_{H;\theta}(h) \approx (h/e^{\mu_H})^{1/\sigma_H},$$

for μ_H large enough. Notice that this means that the relative scaling factor is independent of μ_H, and so, the weighted edf will not depend on μ_H at all. However, since the convergence to the approximation above will be slower for high losses, when increasing μ_H, the scaling at the highest quantiles will be affected long

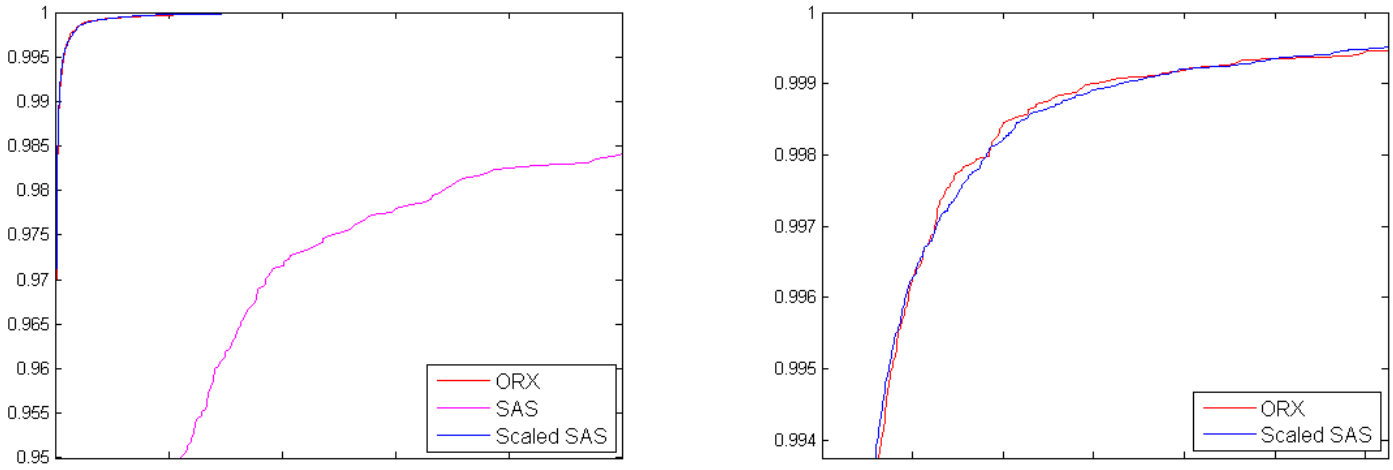


Figure 4 : *The cdfs for the ORX and the weighted and unweighted SAS losses (left) exceeding* €120,000*, and a zoom of the first two cdfs (right).*

after the bulk of the weighted distribution has stabilized. For this reason, we do not recommend that the estimation of the parameters μ_H and σ_H is based solely on $\text{Err}(\mu_H, \sigma_H)$. Instead, after identifying a range of parameter values where $\text{Err}(\mu_H, \sigma_H)$ can be said to have converged, you can use either graphical tools or apply the Kolmogorov-Smirnov test above some chosen high quantile value. In our implementation, we started out by finding the σ_H which minimized the error function over some range of μ_H-values. μ_H was then increased until graphical evaluations implied a good fit. The Q-Q plot of the scaled SAS data versus the ORX data and the resulting cdfs of the complete samples can be seen in Figures 3 and 4 respectively.

There are a number of reasons for not performing any cell specific weighting of losses, or any weighting of ORX losses, even though the assumption regarding the similarity of the ORX and internal data might not have as much support on cell level. First of all, many cells do not have enough losses to give any reliable or robust estimates of scaling factors. Secondly, using small data sets increases the risk of overfitting the data, in which case we would lose the actual information that the ORX and the SAS losses are intended to give in the first place. It should also be clear that the overall error margin will increase if you decide to perform a weighting also of the ORX losses, or of different parts of the data sets in separate procedures.

3.3 Data Mixing

The ORX database requires its members to report all losses exceeding €20,000, while it is optional to report losses below this threshold. As a result, the internal and the ORX losses were first filtered so as to dismiss all losses below €20,000. This is in agreement with paragraph 673, article 2 in Basel Committee on Banking Supervision (2006). If one would like to lower the threshold, to say €10,000, this can be done by first analyzing only the internal loss data to estimate the probability of a loss exceeding the initial threshold at €20,000. The lower part of the distribution can then be estimated using only internal losses while the upper part is estimated by standard means. It is, however, reasonable to doubt whether the available dataset of internal losses is large enough to allow for any reliable estimate of the probability of a loss exceeding the initial threshold *or* of the severity distribution below this threshold.

Because of the lower registration limit of the SAS data, it cannot be mixed with the ORX data directly. Instead, one possibility would be to use the method explained in Aue and Kalkbrener (2006). In

this case, the estimated cdf of the severity distribution will be

$$\hat{F}(x) = \begin{cases} \hat{F}_{ORX}(x) & , x < u' \\ 1 - (1 - \hat{F}_{ORX}(100{,}000))(w_1(1 - \hat{F}_{ORX}^{(100{,}000)}(x)) + w_2(1 - \hat{F}_{SAS}(x))) & , x \geq u', \end{cases}$$

where $F_X^{(u)}(x) = \frac{F_X(x) - F_X(u)}{\bar{F}_X(u)}$ for $x \geq u$, $w_1 + w_2 = 1$, and u' is the current value of \$100,000 in euro. This is straightforward to implement in the POT method, and if the threshold is estimated to be below \$100,000, theory would still allow us to fit a GPD to the SAS-data with a threshold of \$100,000. However, to be able to benefit from the simple approximations of the VaR for aggregated distributions covered in section 2.5, we will instead simply add the SAS data exceeding the threshold to the ORX data when estimating the GPD distribution. The upper limit on the number of extreme values will be chosen so that the threshold does not fall below €120,000 (see the motivation in the previous section) for most cells. This will not restrict the threshold estimates in any substantial way since a threshold of €120,000 would imply far more extreme values than given by standard threshold estimates only considering the number of available losses. Still, one disadvantage of this method is that we will be forced to discard all SAS losses below the chosen threshold.

4 Procedure

Many figures in this chapter and the next will use losses from two cells named *cell 1* and *cell 2*. These cells consist of 108 and 2283 ORX losses respectively. All qualitative conclusions holds for all cells if nothing else is stated. All VaR estimates are with respect to the 99.9th percentile of the yearly loss distribution (as given by paragraph 667 in International Convergence of Capital Measurement and Capital Standards).

4.1 Estimation of Frequency Distributions

The intensities of the Poisson processes were estimated using the standard ML estimate, i.e. by dividing the number of internal events in each cell with the time period since the registration started. The estimates will due to disclosure agreements not be displayed together with their respective cells.

4.2 Threshold Estimation

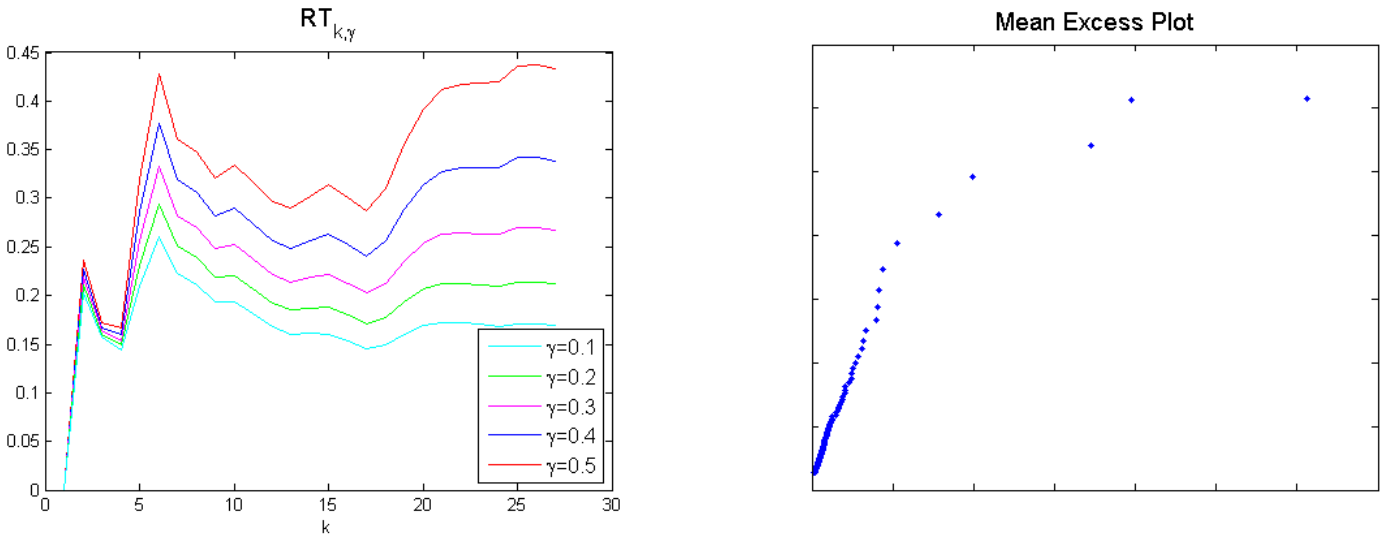


Figure 5 : $\mathrm{RT}_{k,\gamma}$ *as a function of k (left) and the mean excess plot (right). Both figures uses the sample of cell 1.*

Since the SAS-losses only consists of losses exceeding \$100,000, only the ORX-losses were considered when estimating the threshold. The Huisman and the Riess-Thomas methods were implemented in a straightforward manner as described in section 2.7. The parameter γ in the Riess-Thomas method was fixed at 0.5, that is, the estimator was allowed to put a large emphasis on the smallest extreme values. This seemed to accentuate the minima of $\mathrm{RT}_{k,\gamma}$ the most (see Figure 5 (left)). Riess-Thomas was implemented using the Hill estimator, since the estimator has been shown to posses those properties that Riess-Thomas tries to utilize (see section 2.6.1).

Since the mean and median excess functions are normally used graphically, or merely as a validation of an already estimated threshold, we will have to find some method to measure the linearity of the excess functions. Although a subjective graphical approach might be more flexible, the motivation of a particular choice is often unclear (notice, for instance, how difficult it is to decide on a threshold from the mean excess plot in Figure 5 (right)), and furthermore, the method does not allow any meaningful study of robustness. Therefore, we performed an unweighted linear regression on the excess function for each possible threshold, which was followed by an evaluation of the standard information criterions BIC and AIC. The number of extreme losses were then chosen as the k which minimized the respective

criterias. We also evaluated a threshold estimator which, given the ML estimator, minimizes the upper tail Anderson-Darling statistic applied to the severity distribution over all thresholds.

The estimated thresholds were never allowed to indicate less than $5 + N_{losses}/50$ or more than $10 + N_{losses}/7$ extreme observations, where both N_{losses}, i.e. the total number of observations in the (possibly aggregated) cell, and the extreme observations refers to losses from ORX. The threshold value was always chosen as the mean value of the minimal loss included among the extreme losses, and the maximal excluded loss.

4.3 Calibration of Parameter Estimators

The number of losses used for the Hill estimator, Pickands estimator and Huisman's estimator were chosen as $12+n/25$ (or all extreme losses if less were available). This means that the estimators primarily will consider losses roughly corresponding to those severity quantiles that determine the VaR-estimate in section 2.5. The parameter k in the kMedMad estimator was chosen as 8 after a preliminary evaluation using graphical tools and test statistics.

We will in addition to the estimators described in section 2.6 consider an estimator (denoted MoMom-Q) which combines the MoMom-estimate of ξ with a β-estimate based on a fit of some high empirical quantile. Motivated by equation (1), the chosen quantile will be close to $1-\alpha/\mathrm{E[N]}$. To make the estimator more robust with regard to losses exceeding all previously observed losses, the index of the loss which is "matched" against the theoretical quantile will never be allowed to be smaller than 5, using the notation of section 2.6. To more precise, $\hat{\beta}$ will be estimated using

$$\frac{n+1-\tilde{n}}{n} = 1 - \left(1 + \hat{\xi}_{PWM}\frac{x_{\tilde{n}}}{\hat{\beta}}\right)^{-1/\hat{\xi}},$$

where $\tilde{n} = \max\{\lceil n\alpha/\mathrm{E}[N]\rceil, 5\}$.

All the SAS losses are weighted in the parameter estimates, using the weights $w_i = 1/\mathrm{P}(H < X^i_{SAS})$ according to section 3.2. In the ML estimates, the weights come in as exponents in the factors making up $l(\xi, \beta)$, which means that the terms in the sum in equation (4) will be weighted by multiplication of w_i. The rationale for this can be understood by acknowledging that a weighting of a fictitious sample $\{X_1, X_2\}$ by $w_1 = 2/3, w_2 = 4/3$, implies that it will be handled as the unweighted sample $\{X_1, X_2, X_2\}$.

4.4 Approximations of Risk Measures and Convergence of Monte Carlo Estimates

Two types of analyses were performed to test the risk measure approximations in section 2.5. First, the approximations from section 2.5 as well as MC estimates of the VaR and ES were calculated for all cells with more than 50 ORX losses and at least one internal loss (33 cells in total). This was done using both ML and MoMom-Q estimates.

Secondly, the risk measure estimates were compared by using the edf in the body of cell 1 while varying $\hat{\xi}$ and holding $\hat{\beta}$ constant in the tail distribution. The losses exceeding the threshold were thus ignored in these cases. Since $\hat{\xi}$ was allowed to vary between -1 and 5, $\hat{\beta}$ was chosen as $3/2 * (\mathrm{Loss}_{max} - u)$ with obvious notation. Notice that if $\hat{\beta}$ is chosen as too low a value, the empirically estimated part of the distribution might have an unreasonably large impact on the simulations when $\hat{\xi}$ is negative, in which case the largest possible simulated loss is $u - \hat{\beta}/\hat{\xi}$. All MC estimates were calculated using 50 million simulations of the aggregated loss distribution.

Besides comparing the approximations with the MC estimates, the standard deviations (SD) of the

MC estimates were also estimated. This was done by dividing the unordered 50 million simulations into 50 buckets and computing the sample SD of estimates from the different buckets. The relative error was then defined as the estimated SD divided by the estimated value of the risk measure using all 50 million losses. The threshold was chosen as the empirical 0.9-quantile in all these studies.

No significant effect was noted on the precision of the approximate expressions or on the rate of convergence of the MC estimates with respect to the β-parameter (excluding non-relevant degenerate cases) and the threshold. Note that the threshold in all practical cases will be chosen so that the important severity quantile (see equation (1)) belongs to the GPD-tail.

4.5 Simulation of Data Loss

There is a natural data shortage in all cells since ORX only has a limited number of members, all contributing data from a short period of time. To study the effects of this, we will simulate data loss by drawing losses without repetition from a relatively large data set to form a smaller data set. This is done a multiple number of times, and then the estimated relative error (sample standard deviation divided by the value of the risk measure when using the complete data set) of the VaR can be calculated by using the estimates from the smaller data sets. The VaR will be calculated using the approximation in equation (3).

The larger data sets were chosen as the ORX cells with more than 2,000 losses and a nonzero number of internal losses. 10,000 smaller data sets (each consisting of 200 losses) were constructed from each of these sets. The threshold was set at the empirical 0.9-quantile in all cases.

5 Results

This section presents the results and analysis of the tests that were performed in this study, some of which were described more closely in the previous section. First, we analyze the choice of severity distribution and decide to continue with POT method. We then move on to study the different parameter estimators for the GPD, and decide to reject the estimators of Hill, Huisman and Pickands. Next, the risk measure approximations described in section 2.5 are studied together with the corresponding MC estimates, and the stability of the parameter estimators are examined with regard to both data loss and the chosen threshold. A cell aggregation is then motivated by using both qualitative and quantitative arguments. Finally, this aggregation is combined with the MoMom-Q estimator to numerically estimate the correlation bounds given in section 2.13.

5.1 Analysis of Severity Distributions

After fixing the threshold at the 0.85-quantile and estimating the GPD, as well as the Weibull and the lognormal distribution by using ML estimates, the severity cdfs were plotted as seen in Figure 6.

The plots indicates that the GPD is the most suitable among the fitted distributions (which is the same conclusion that was drawn in Aue and Kalkbrenner (2006)). This is also confirmed by Figures 7, 8 and 9 showing the Q-Q plots of the same distributions. You should note that the Q-Q plots for the GPD only considered the extreme losses, and that the corresponding tick marks in each plot represent the same values for both axes. Further arguments for the use of POT methods in operational risk modeling can be found in Embrechts et al. (2006).

As can be seen from the plots, the lognormal and the Weibull distributions are not able to capture the characteristics of the most extreme observations, which becomes especially evident when the number of considered losses increases. Estimators which fits the tail of some data set to a lognormal or Weibull distribution have been proposed, but will not be considered in this thesis. This can be motivated by the reasonable fit of the GPD in the plots above, and also by the second theorem in EVT. Furthermore, as is illustrated in section 5.3, the body of the distribution has in most standard cases a limited impact on the regulatory capital, which justifies our use of the edf in the body of the severity distribution.

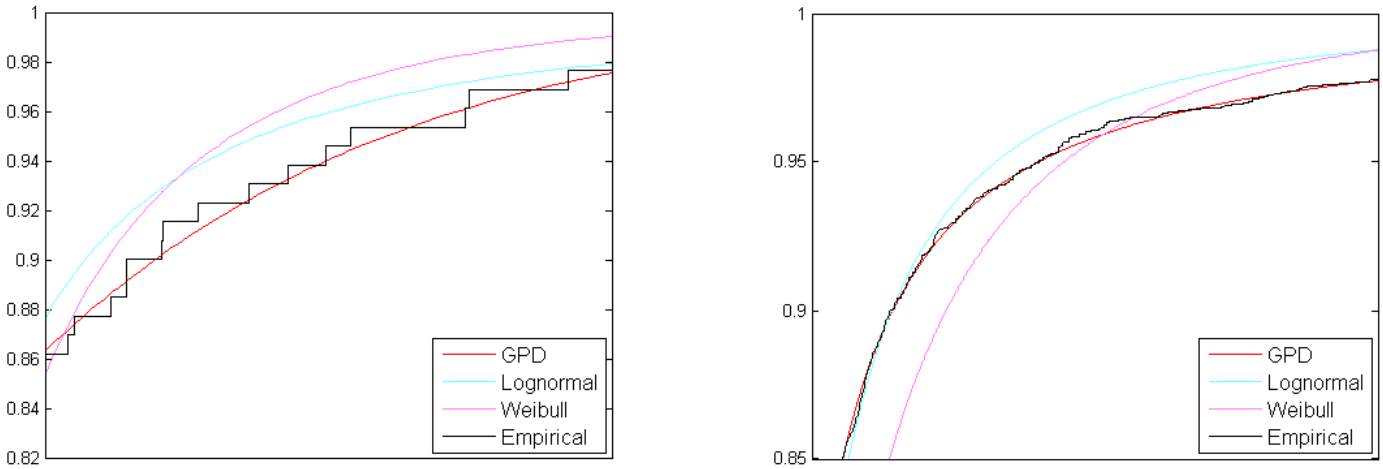


Figure 6 : *Estimated and empirical cdfs for cell 1 (left) and cell 2 (right).*

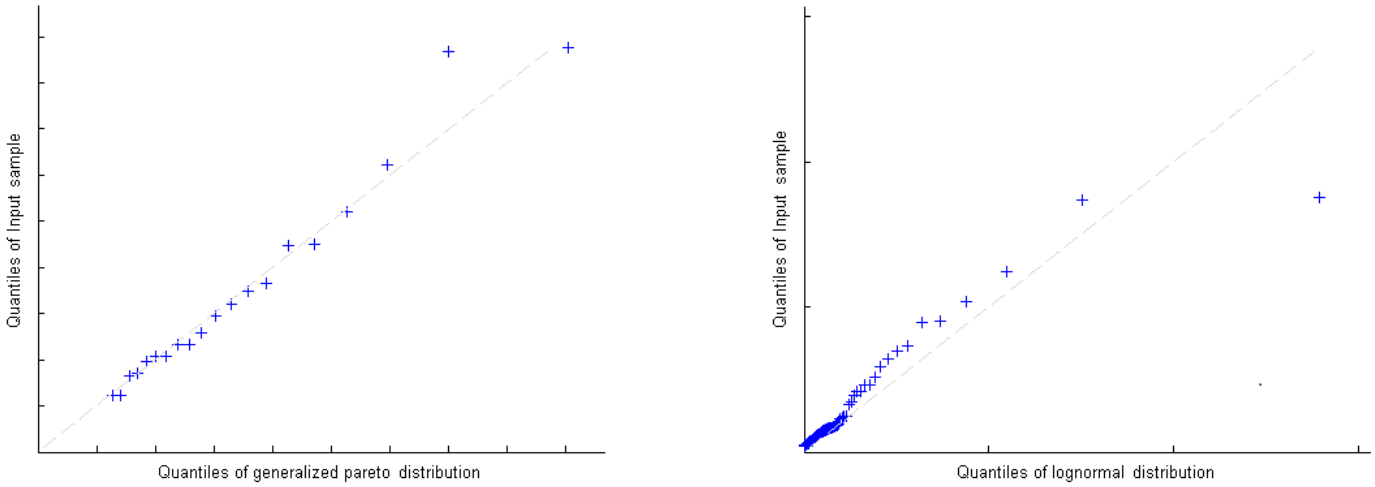


Figure 7 : *Q-Q plots based on an estimated GPD (left) and lognormal (right) distribution for cell 1.*

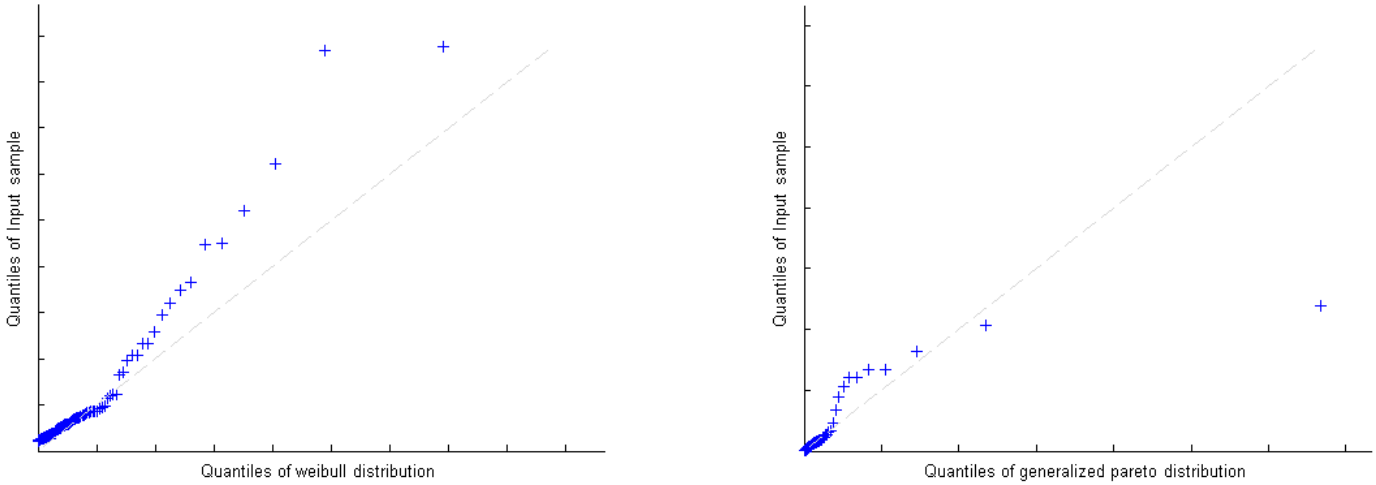


Figure 8 : *Q-Q plots based on an estimated Weibull (left) distribution for cell 1 and an estimated GPD (right) for cell 2.*

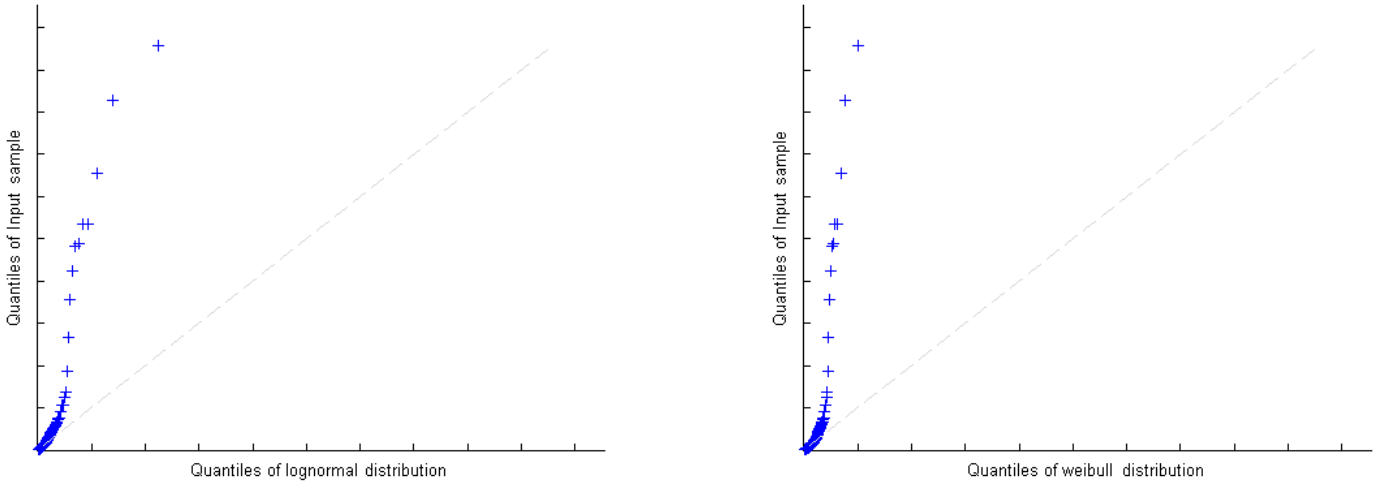


Figure 9 : *Q-Q plots based on an estimated lognormal (left) and Weibull (right) for cell 2.*

In Figure 10 you can see the distribution of the exceedances among the quantiles of the estimated GPD for the two cells. The quantiles clearly seems to stem from a uniform distribution without any apparent bias, which gives further weight to a GPD-assumption.

Incorporating the SAS losses has the effect of smoothing out the edfs and providing much needed observations from the tail of the distributions. This is illustrated in Figure 11 where the empirical cdfs as well as the estimated GPDs are shown when using the ORX losses, the SAS losses, and all losses from both sources. The plots exemplifies the contribution of the SAS losses when the extreme SAS losses outnumbers the extreme ORX losses (left), and when the number of extreme losses coming from the two data sets are fairly equal in quantity (right).

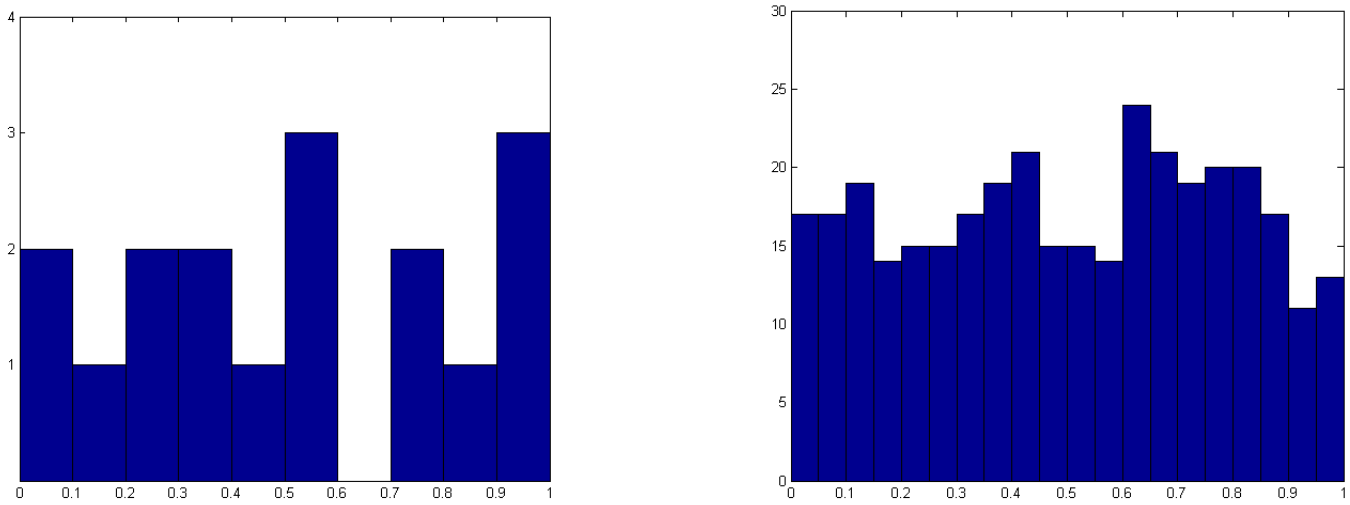


Figure 10 : *Histograms showing quantile partition of losses from cell 1 (left) and cell 2 (right) with respect to the estimated GPDs.*

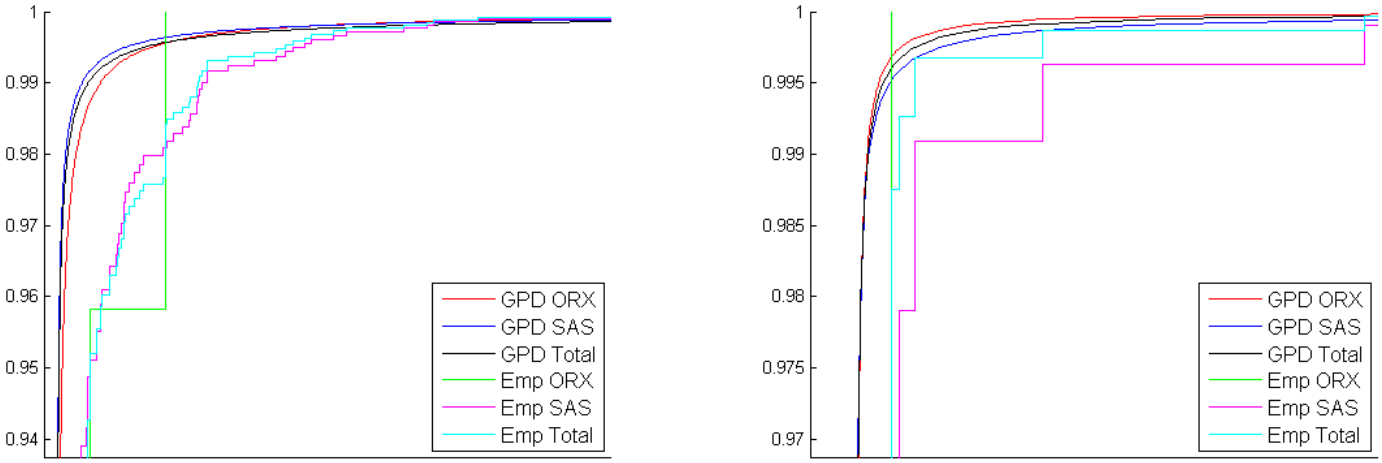


Figure 11 : *Empirical distributions and estimated GPDs when using ORX, SAS and combined data.*

5.2 Analysis of Parameter Estimators

Fixing the threshold at the empirical 0.9-quantile and applying all estimators covered in section 2.6 (as well as MoMom-Q) to the losses in cell 1 and cell 2 gave the cdfs displayed in Figures 12 and 13. Especially notice how the MoMom-Q estimator sacrifices a close fit of the larger part of the edf to obtain an excellent fit of the highest empirical quantiles (see Figure 13 (right)), which all other estimators underestimates. It should be noted that the fits obtained when applying the Hill and Huisman estimators to cell 1 were exceptionally poor in comparison to other cells, even though both estimators proved to be very sensitive and unreliable with regard to the chosen threshold.

Figure 14, 15 and 16 shows the UTAD and the KS p-value respectively as functions of the number of available losses, when applying all the parameter estimators to all cells with at least one internal loss and more than 50 ORX losses.

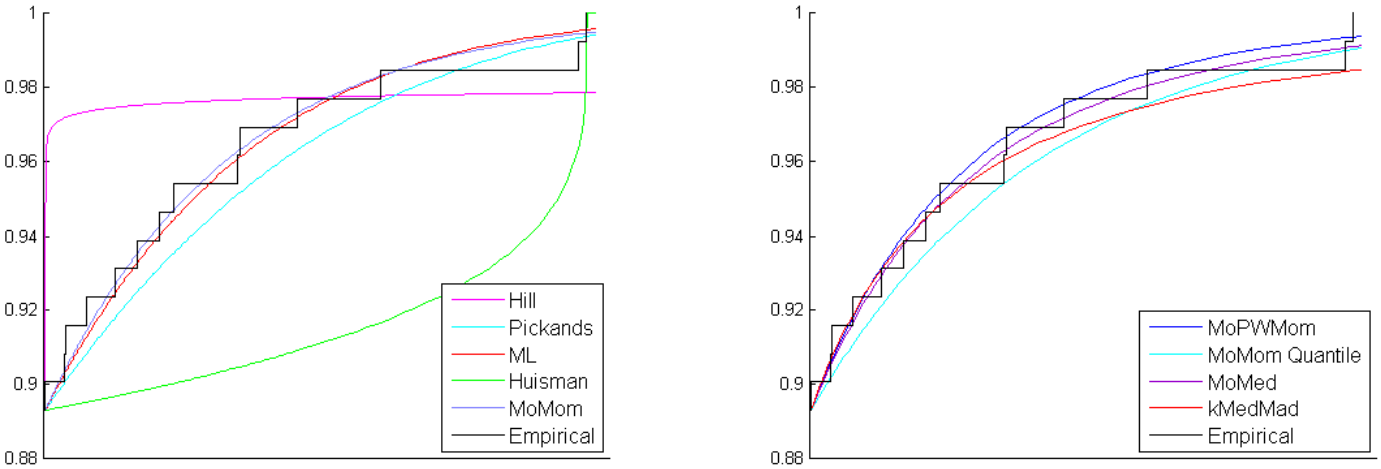


Figure 12 : *Estimated and empirical cdfs for cell 1.*

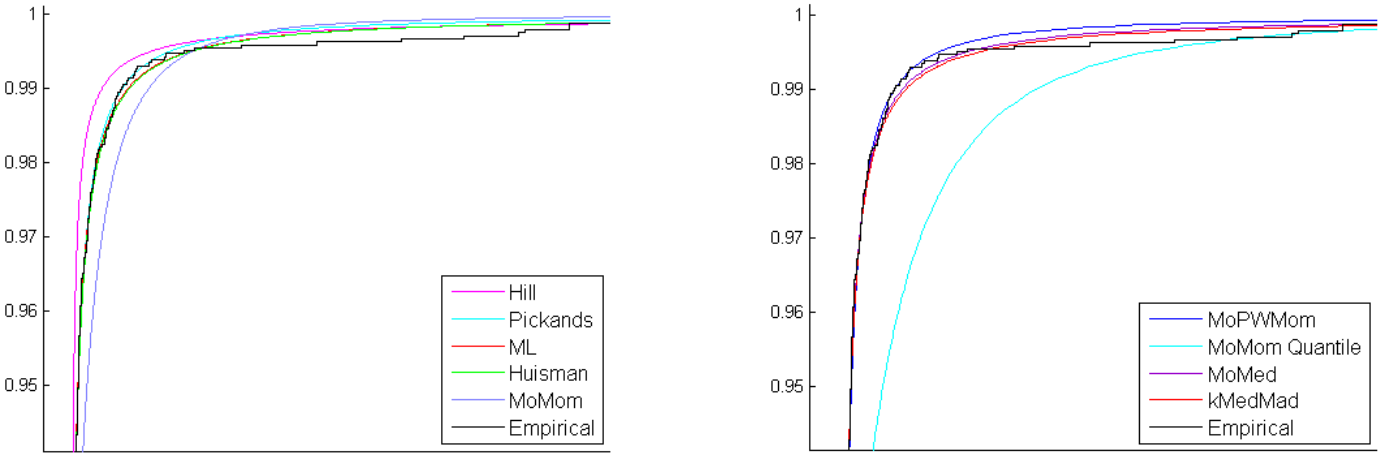


Figure 13 : *Estimated and empirical cdfs for cell 2.*

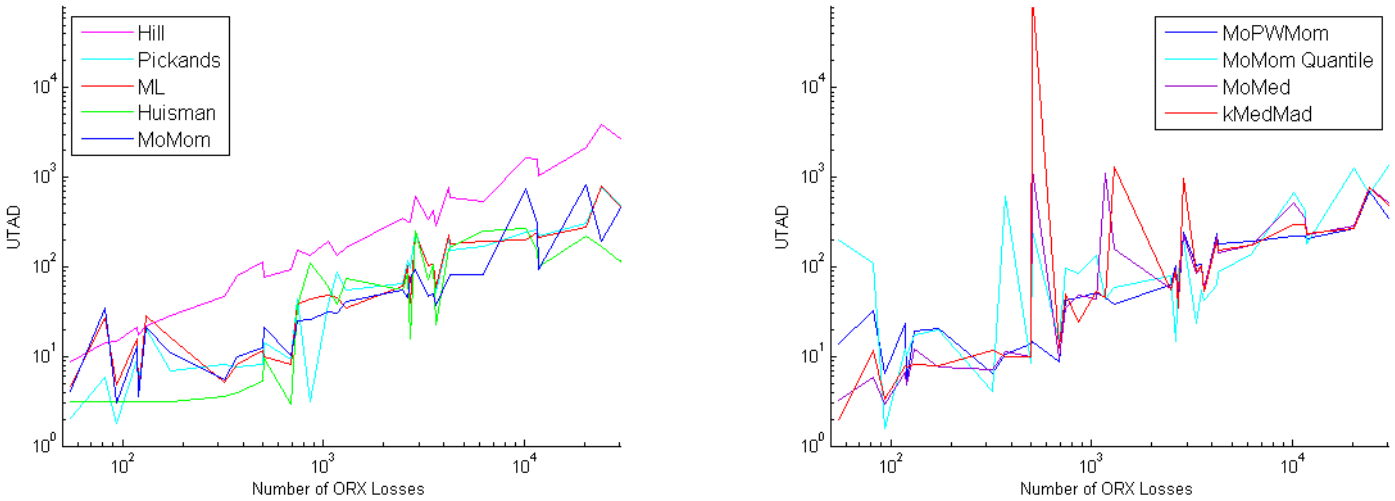


Figure 14 : *UTAD associated with the estimated GPD as a function of the number of available ORX losses.*

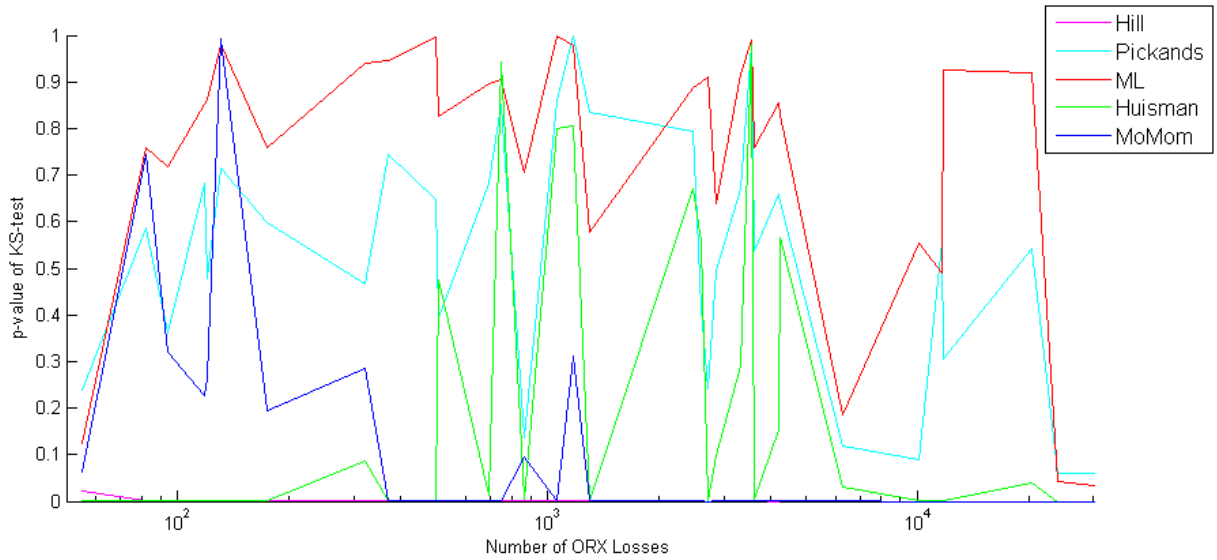


Figure 15 : *KS p-value associated with the estimated GPD as a function of the number of available ORX losses.*

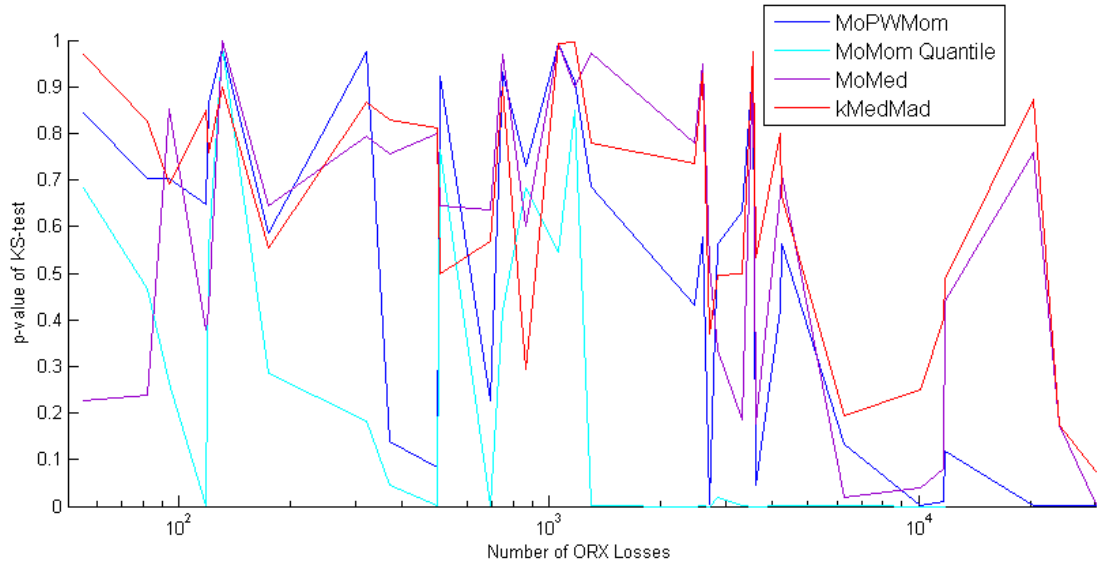


Figure 16 : *KS p-value associated with the estimated GPD as a function of the number of available ORX losses.*

MoMom and MoMom-Q performs decently when applied to smaller data sets according to the KS p-value, but for larger sets have p-values practically equal to zero. The reason for this behaviour is most likely that the data is generally more heavy-tailed for larger data sets, since smaller data sets often lack the most extreme losses. Since MoMom cannot give $\hat{\xi} > 1/2$, a poor overall fit is natural for data with very heavy tails. What is more important to notice though, is that the two estimators according to UTAD are comparable to the best performing estimators as the number of observations grows. This supports the notion that a ξ-estimate which is biased towards lower values (or limited as when using the MoPWMom or MoMom estimates) often gives more precise estimates of the upper tail. This has previously been discussed in McNeil and Saladin (1997) which concludes that the ML estimate often overestimates the highest quantiles, something that is also supported by Galfond (1982), which suggests that the estimate of the shape parameter is truncated at 1/2 or 1.

Even though the Huisman estimator seems to perform very well in the upper tail when applied to smaller data sets, this is in most cases due to a negative ξ-estimate (problems with this is discussed in section 5.5) which is brought about by an insufficient number of available extreme values. We should also mention that the Huisman and the Pickands estimator at times proposed parameter estimates which gave $F_{sev}(x) = 1$ for x smaller than the highest observed loss, which is obviously unacceptable. These two estimators will for this reason be discarded from further analysis, while the Hill estimator is rejected based on its weak performance (see Figure 14 (left)).

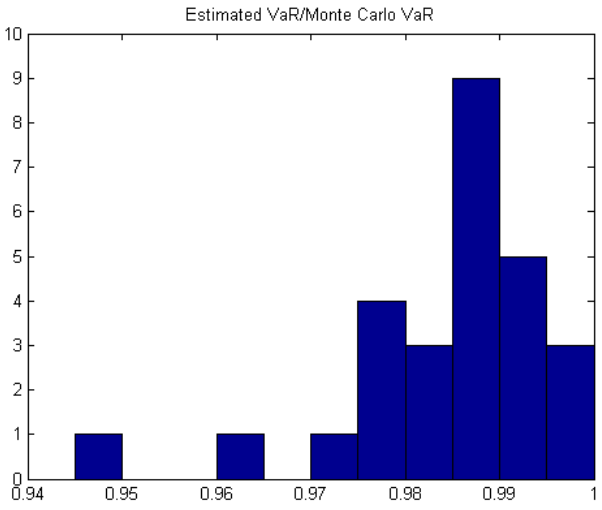


Figure 17 : *Histogram of Estimated VaR/Monte Carlo VaR using ML estimates.*

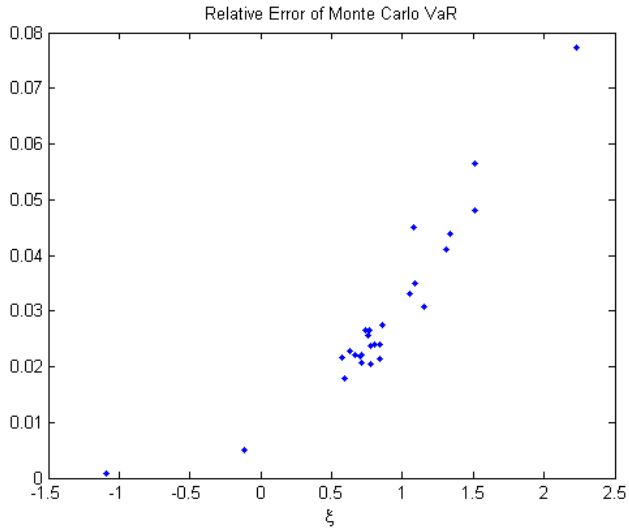
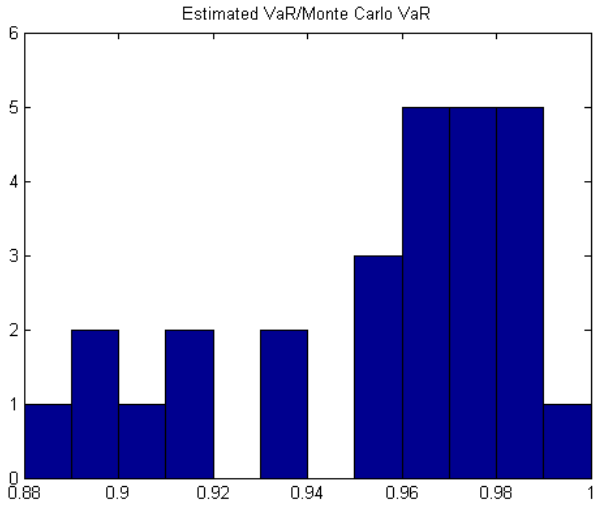


Figure 18 : *Relative Error of the Monte Carlo VaR as a function of the estimated shape parameter when using ML estimates.*

5.3 Approximations of Risk Measures

This section present the results of the analysis described in the first two paragraphs of section 4.4.

As you can see from the histogram in Figure 17, the approximation in equation (3) seems to be fairly good when using the ML estimator on real loss data. Furthermore, Figure 18 indicates that the relative errors of the MC VaRs are dependent on the value of the shape parameter, something which is investigated more carefully in the next section. The results from the ES-approximation were left out in this case since ξ is estimated to be larger than one for many cells.



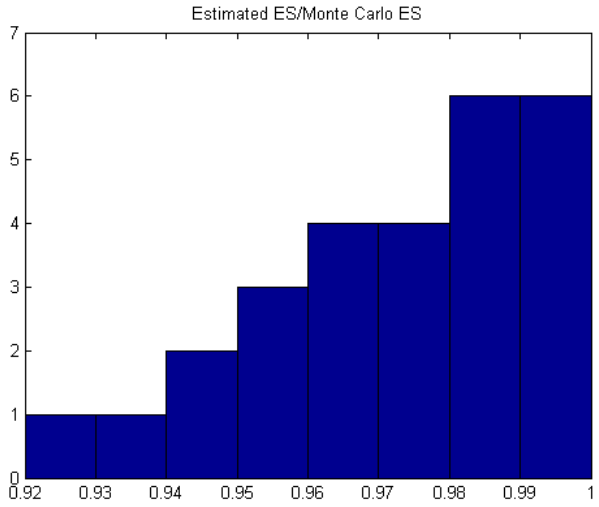


Figure 19 : *Estimated VaR/Monte Carlo VaR (left) and Estimated ES/Monte Carlo ES (right) using MoMom-Q estimates.*

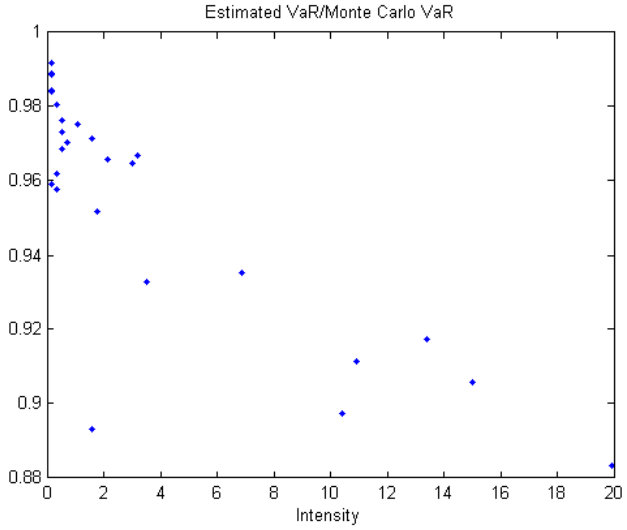


Figure 20 : *Estimated VaR/Monte Carlo VaR as a function of the Poisson intensity when using MoMom-Q estimates.*

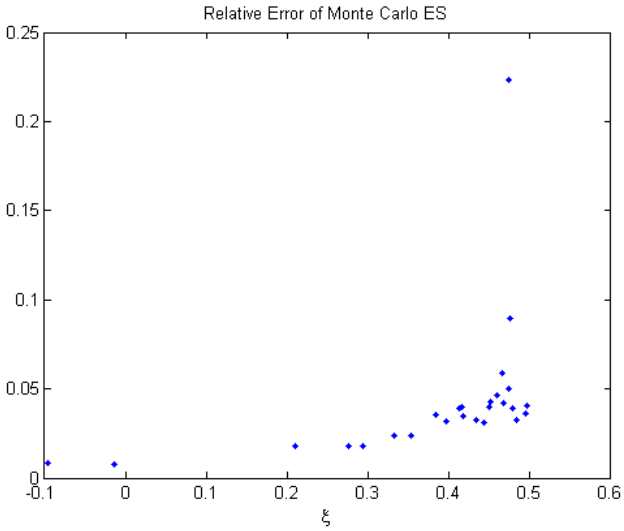
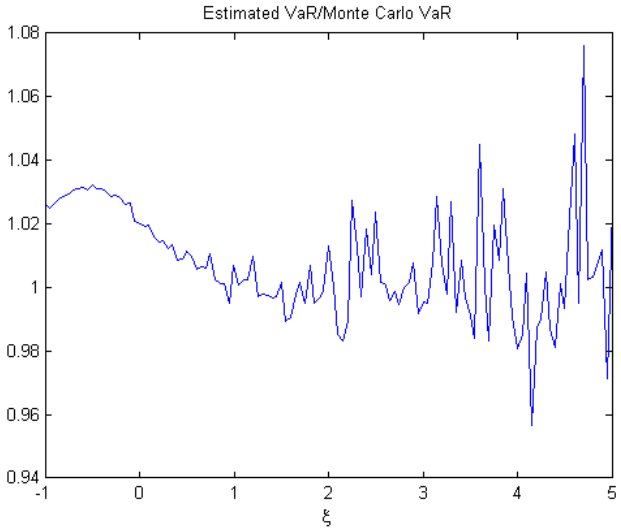


Figure 21 : *Relative Error of Monte Carlo ES as a function of the estimated shape parameter when using MoMom-Q estimates.*



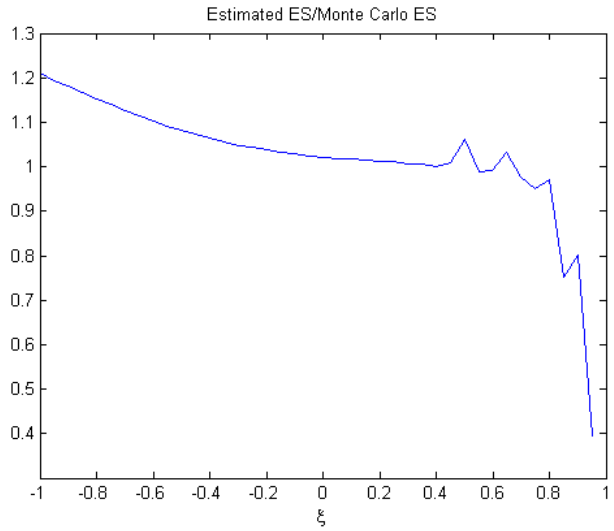


Figure 22 *VaR (left) and ES (right) approximations divided by the respective Monte Carlo estimates as a function of the shape parameter.*

The exact same simulations were made using MoMom-Q estimates, and the results can be seen in Figure 19. The performance of the VaR approximation when using this estimator is further illuminated in Figure 20, were it is seen that the underestimates are correlated with large intensities of the associated Poisson distribution. This can be intuitively motivated by the fact that the convergence to the "sub-exponential property" (which is the limit in which the approximation holds) of the severity variable, X , i.e. that

$$\lim_{s\to\infty} \frac{\mathrm{P}(X_1 + ... + X_n > s)}{\mathrm{P}(X > s)} = n, \tag{16}$$

where $X_1, ..., X_n$ are independent copies of X, will be slower in s for larger n and smaller ξ. This is also supported by the fact that the ES approximations are somewhat better since they are approximations of larger quantiles where the convergence is faster. Here, the aggregated loss should be thought of as $S_n = X_1 + ... + X_n$, where n in practice is Poisson distributed, and obviously in general will take on larger values as the intensity increases. The analysis is further complicated by the fact that s will vary with both the intensity and the shape parameter. As expected, Figure 21 suggests that the relative error of the MC ES also increases together with the shape parameter. As a side note, we will mention that simulations where the intensity was "controlled" and allowed to vary when using MoMom-Q estimates confirmed that the VaR approximation gives more pronounced underestimates as the intensity is increased.

By varying the shape parameter of the tail distribution (as described in the second paragraph of section 4.4), we can further investigate the behaviour of the risk measure approximations as functions of ξ. Notice that Figure 22 indicates that the approximation is very useful also for larger values of ξ, although it is computationally expensive to validate its precision in the presence of heavy tails.

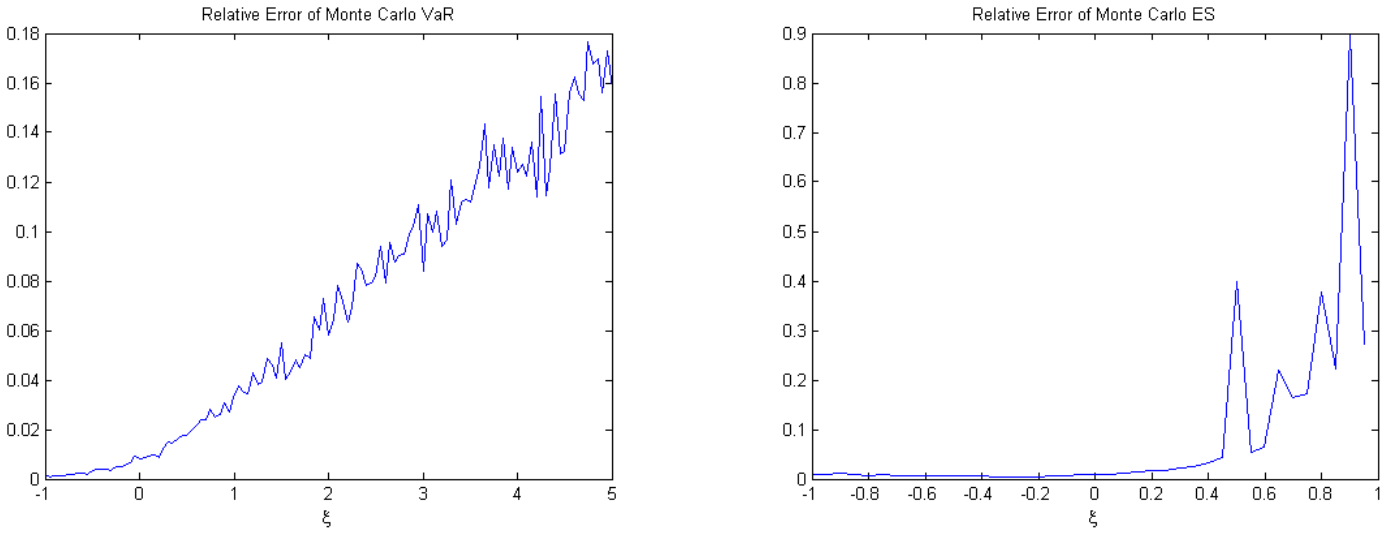


Figure 23 : *Relative errors of the Monte Carlo estimated VaR (left) and ES (right) as a function of the shape parameter. The estimation procedure and the data is described in section 4.4.*

5.4 Convergence of the Monte Carlo Estimates

Figure 23 reveals the slow convergence of the MC estimates for larger values of ξ. Remember, however, that all relative errors have been estimated from risk measure estimates using one million simulations each, while the risk measure estimates in the previous section were calculated using fifty million simulations. Further, notice that the relative error of the estimated VaR seems to be linear in ξ. Denoting the number

of simulations by N_{sim}, this can be theoretically motivated by the fact that

$$\hat{\text{VaR}}_\alpha(S) = \hat{F}^{-1}_{agg}(1-\alpha) \sim \text{N}\left(F^{-1}_{agg}(1-\alpha), \frac{1}{N_{sim}} \frac{\alpha(1-\alpha)}{(f_{agg}(F^{-1}_{agg}(1-\alpha)))^2}\right), \text{ as } N_{sim} \to \infty, \tag{17}$$

(see Hult and Svensson (2009)) under suitable conditions, when the edf is used to estimate the VaR. By using the not so far-fetched assumption that S is heavy-tailed, i.e. that $F_{agg}(x) \approx 1 - ax^{-1/\xi'}$ as $x \to \infty$, this gives

$$\text{Rel Err}(\hat{\text{VaR}}_\alpha(S)) = \frac{\text{SD}(\hat{\text{VaR}}_\alpha(S))}{\text{VaR}_\alpha(S)} \approx \frac{\xi'}{\sqrt{N_{sim}}}\sqrt{\frac{1-\alpha}{\alpha}}, \text{ as } N_{sim} \to \infty. \tag{18}$$

The approximation $\xi \approx \xi'$ can then be motivated by $\bar{F}_{agg}(s) \approx \bar{F}_{sev}(s)\lambda$ for sufficiently large s (see equation (16)), which gives

$$\lim_{t\to\infty} \frac{\bar{F}_{agg}(ts)}{\bar{F}_{agg}(t)} = \lim_{t\to\infty} \frac{\bar{F}_{sev}(ts)}{\bar{F}_{sev}(t)} = s^{-1/\xi},$$

and therefore explains the roughly linear relationship in Figure 23 (left).

5.5 Quantile Estimation

This section demonstrates the effect that different threshold and parameter estimators have on the quantiles of the estimated severity distributions. According to the results in section 5.3, the highest quantile estimates are to a large extent what ultimately determines the VaR-estimate of the cell in question, as well as the total capital requirement when assuming perfect correlation between cells. We will moreover also highlight the importance of data sufficiency, and how ignoring this aspect can result in very inconsistent estimates. We do *not* claim to be able to evaluate the actual values of the estimates by solely using loss data. While it might be justified to study relative characteristics of different estimators, a fair risk assessment must always include some type of scenario analysis or validity judgement independent of the external data. The axis values have been normalized in all quantile and threshold plots.

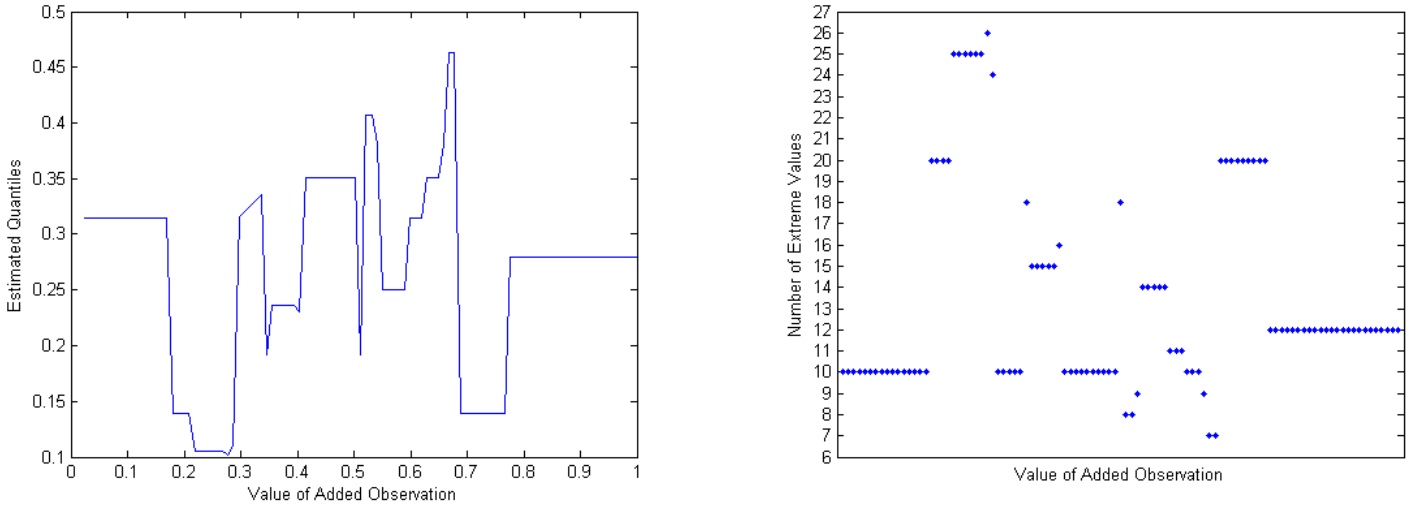


Figure 24 : *The estimated threshold (left) and the number of suggested extreme values (right) as a function of one added loss when applying the median excess estimator to cell 1.*

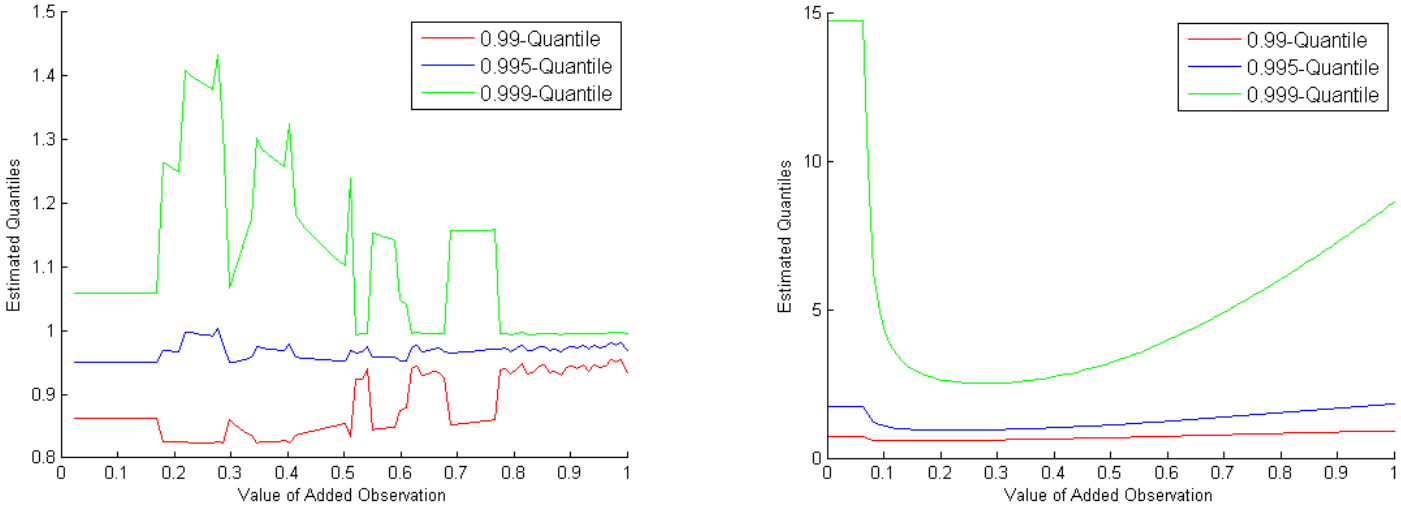


Figure 25 : *The estimated quantiles of the severity distribution (as a function of one added loss) using ML estimates with the median excess threshold applied to cell 1 (left) and a fixed threshold at the* 0.9*-quantile (right) applied to a cell with 94 ORX losses.*

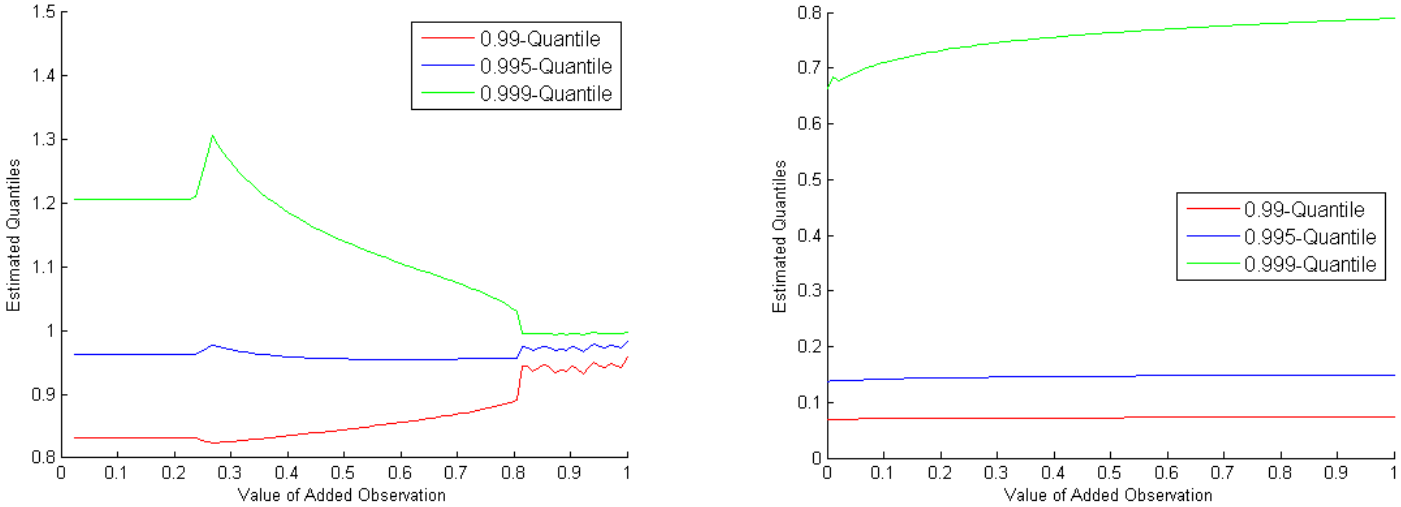


Figure 26 : *The estimated quantiles of the severity distribution (as a function of one added loss) using ML estimates with a fixed threshold at the* 0.9*-quantile and the losses from cell 1 (left) respectively cell 2 (right).*

While conventional threshold estimators differ in both stability properties and average estimated threshold quantile, they are all unstable enough (when handling relatively small data sets) to rule out any robust quantile estimates. The highest threshold estimates are generally given by the estimator minimizing the UTAD statistic, and all estimators are more prone to suggest a lower percentage of extreme values for samples with a larger number of observations. The most unstable estimates are given by the Huisman and the median excess estimator. In Figure 24, you can see the estimated thresholds and the number of estimated extreme values, respectively, when applying the median excess estimator to the sample from cell 1 with one added observation. The plots are drawn as a function of the severity of this added loss (this can be seen as a translation of the SF), which is allowed to vary over 101 equally spaced values ranging between the lowest and the highest observation from the initial sample.

When studying larger data sets, the estimates are more stable. On the other hand, in these cases, the benefit of using any one of the estimators based on loss data is negligible when taking their expensive algorithms into consideration.

As illustrated in Figure 25 (left), the threshold estimators further render equally unstable estimates

of the severity quantiles. With the threshold at a constant empirical quantile, the quantile estimates are smoothed out, which is exemplified in Figure 25 (right). However, the figure also displays the irrational behaviour of the estimates which arises since the cell does not contain sufficiently many extreme values, and the added loss therefore have too great an influence on the estimates. The problem diminishes as the number of available losses increases (see Figure 26 (right), in which only a small negligible jump is seen in the quantile estimate as the value of the added loss first exceeds the threshold).

When the number of extreme values are few, another problem often arises. If ξ is estimated to be less than zero, the quantiles behave as shown in Figure 26 (left) or Figure 27. Remember, at this point, no consideration has been taken to the estimated intensity of the associated event variable, why the relevant quantile (the quantile which according to section 5.3 will determine the VaR with very good precision) might lie anywhere within the range of quantiles shown. In other words, the VaR of this cell will increase *or* decrease when adding a new observation depending on how many internal losses we have available. Furthermore, it is doubtful whether it is reasonable that the estimated capital requirement sometimes is lowered when a hypothetical added observation is increased.

The cause of the phenomenon is illuminated by studying the pdf of the estimated GPD (see Figure 28). When a large loss (about equal in size to the previously highest loss) is added, the estimate of ξ is lowered even further in which case the shape of the pdf is characterized by a very abrupt negative jump just before the free variable reach $u - \hat{\beta}/\hat{\xi}$. When a small loss is added to the data set, the pdf maintains its previous shape. The problem is brought on by an insufficient number of extreme losses and can in many cases be avoided by the inclusion of some database containing publicly reported losses (e.g. SAS OpRisk Global Data). There are nonetheless of course cases when extreme losses are unavailable and no suitable cell aggregation can be performed to solve the problem (see section 5.8 for a further discussion). Therefore, the issue should primarily be addressed at the level where the parameter estimates are made.

Obviously, adding the SAS losses generally has a stabilizing effect. Nevertheless, since the probability scaling in section 3.2 was done on the complete set of losses, the final effect of including the SAS data will be very dependent on the chosen cell aggregation.

The quantile estimates of cell 1 respectively cell 2 when using the MoMom-Q estimator are shown in Figure 29. This estimator is not able to solve the problems related to data sufficiency, but displays some useful characteristics when the number of available losses increases. By comparing Figure 29 (right) with Figure 26 (right) you can see that the ML and the MoMom-Q estimators imply stability with respect to different losses, since the use of the latter causes the sensitivy to be concentrated to the empirical quantile giving the approximation in equation (1). The difference is even more pronounced when the added loss is allowed to exceed the current largest loss, in which case the ML estimator allows the quantile estimates to approach infinity much faster than the MoMom-Q estimator. The MoMom and MoPWMom estimators behaves in much the same way as the ML estimator with regard to quantile estimates, and no general conclusion can be drawn regarding the actual quantile values given by different estimators.

The general problem of the conventional threshold estimators is that the different functions that are subject to minimization typically have multiple local minimas, and imply fits of comparable quality for often substantially different thresholds. Added losses can then easily have a very large impact on the number of suggested extreme values without changing the shape of the function being minimized in any significant way.

As a final word of caution, it should be noted that while criticism of conventional threshold estimators is easily motivated in a model assuming perfect correlation between cells and using VaR to determine the final capital requirement, the analysis grows more complicated as the model is refined.

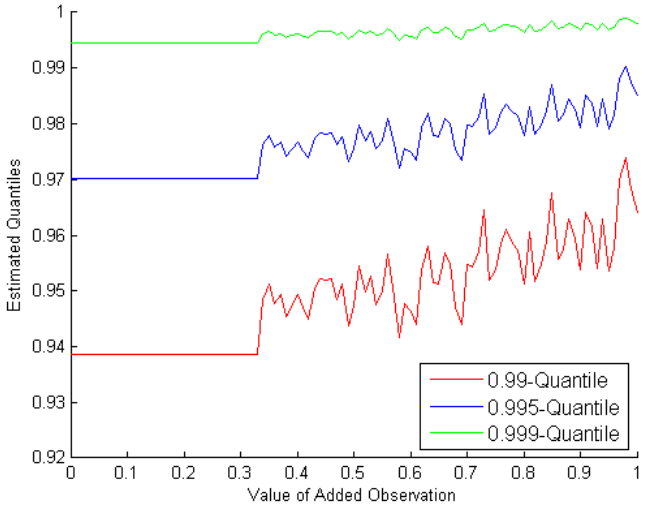


Figure 27 : *The estimated quantiles of the severity distribution (as a function of one added loss) using ML estimates with a fixed threshold at the* 0.9*-quantile and losses from a cell with 84 ORX losses.*

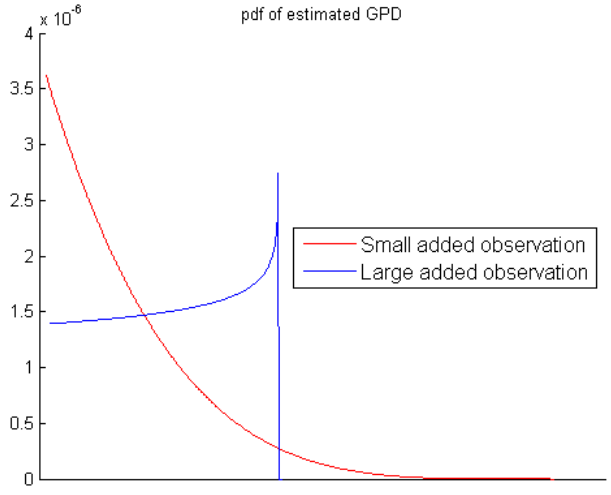


Figure 28 : *The pdf of a GPD estimated from cell 1 after adding one loss with the same value as the smallest and the largest previous extreme value respectively.*

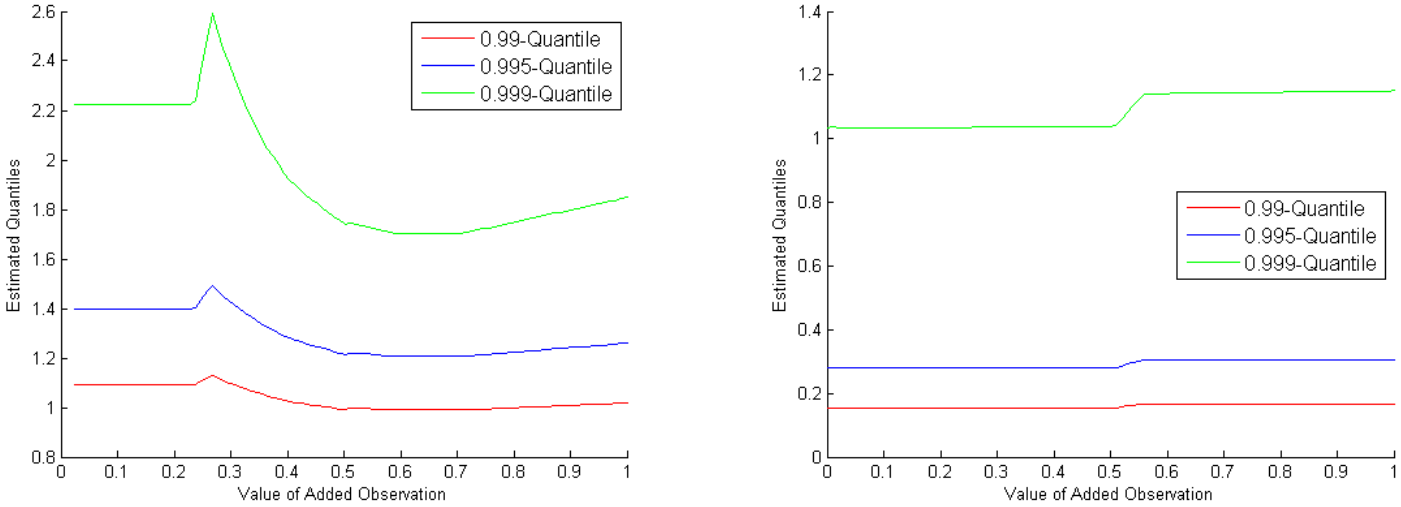


Figure 29 : *The estimated quantiles of the severity distribution (as a function of one added loss) using MoMom-Q estimates with a fixed threshold at the* 0.9*-quantile and the losses from cell 1 (left) respectively cell 2 (right).*

5.6 Simulation of Data Loss

This section presents the results of the analysis described in section 4.5. Table 3 in section 8.12 in the appendix displays the relative errors of the VaR estimates with different parameter estimators divided by the corresponding relative errors when using the MoMom-Q estimator. While the BL-ET specification cannot be publicized due to disclosure agreements, the ML-estimate of the shape parameter is given as an indicator of the "heavy-tailedness" of the cell in question. The table shows that both the ML and the kMedMad estimates are very dependent on which losses that are drawn from the larger sets, while the MoMom-Q estimator gives the most stable estimates. The MoMed estimator was excluded from the study due to its slow convergence, but is expected to perform on the same level as kMedMad since the large relative errors in most cases is associated with occasional overestimates of the shape parameter, which is unbounded from above for all estimators but those based on probability moments.

We will look at an example where the procedure described in section 4.5 is applied to a single cell. All VaR estimates in the histograms have been divided by the corresponding estimate when all losses are included. We then speak of an "overestimation" as if the VaR estimate obtained when including all losses in the estimation was the actual sought value.

Figure 30 shows that the ML estimator is very sensitive with regard to data loss, and often greatly overestimates the VaR (with a factor of about 10^4 in the most extreme case). The peaks, which are most pronounced in Figure 31 (left), come about because the most extreme losses from the cell affect the parameter estimates very strongly in those cases that they are included in the smaller sets. For instance, the rightmost peak in the mentioned figure corresponds to those cases in which the single largest loss from the cell was included. Here, the largest loss was about three times as high as the second largest loss. The effect is not as prominent when the largest losses are more similar in size.

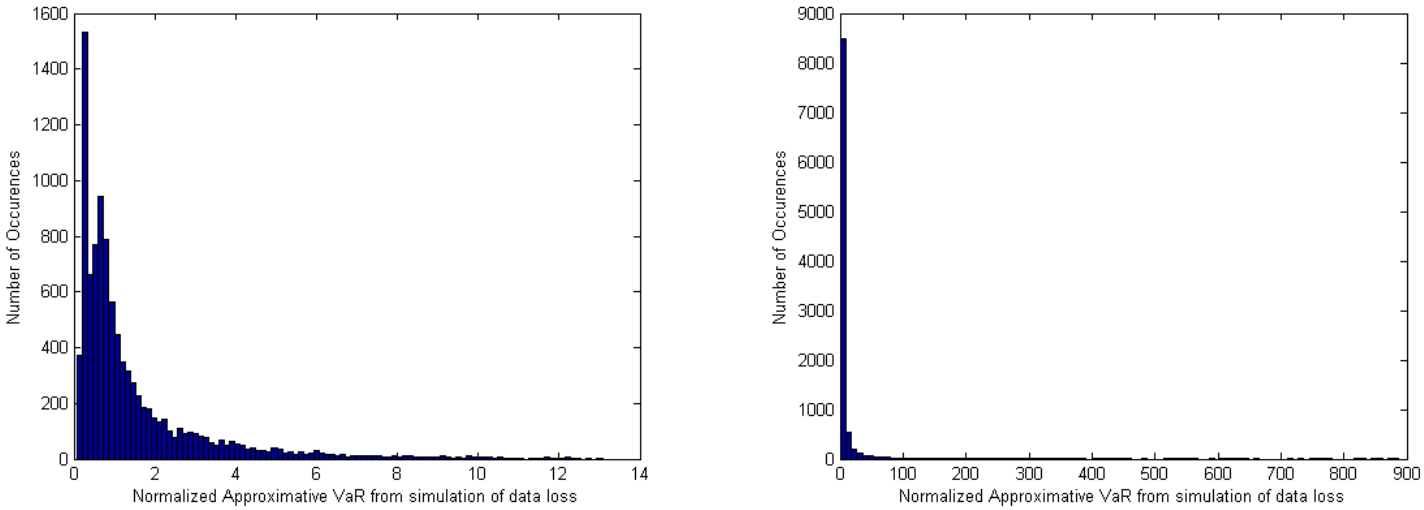


Figure 30 : *Histograms showing the distribution of the normalized approximate VaR of a cell when simulating data loss and using the ML estimator (left) and the kMedMad estimator (right) respectively. The largest percent of the VaR estimates has been excluded.*

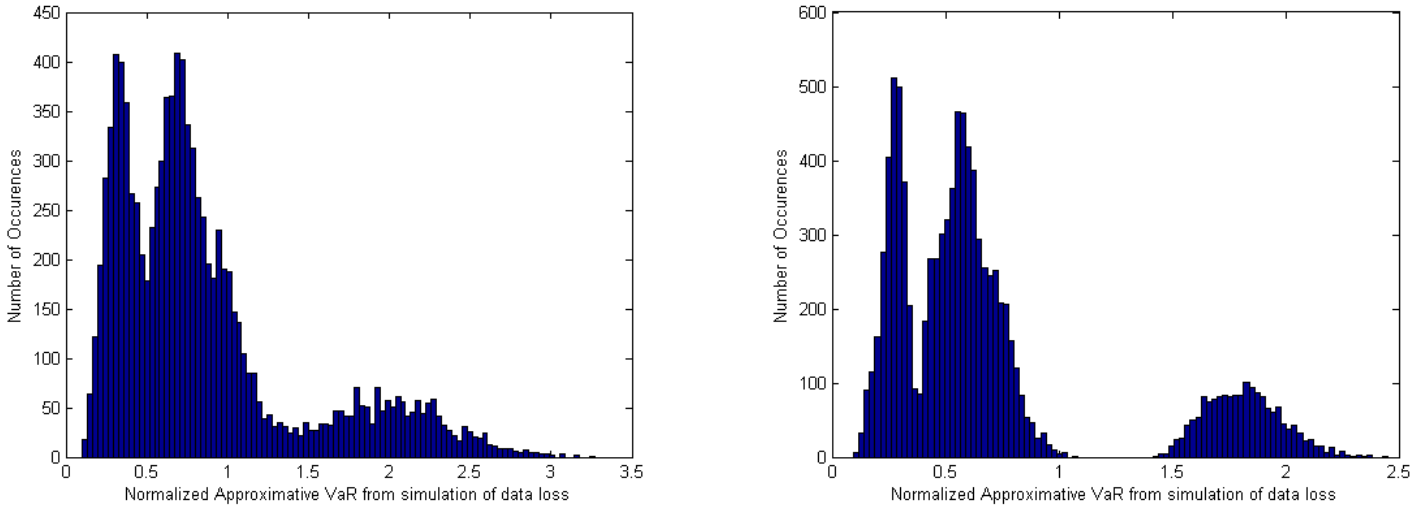


Figure 31 : *Histograms showing the distribution of the normalized approximate VaR of a cell when simulating data loss and using the MoPWMom estimator (left) and the MoMom estimator (right) respectively.*

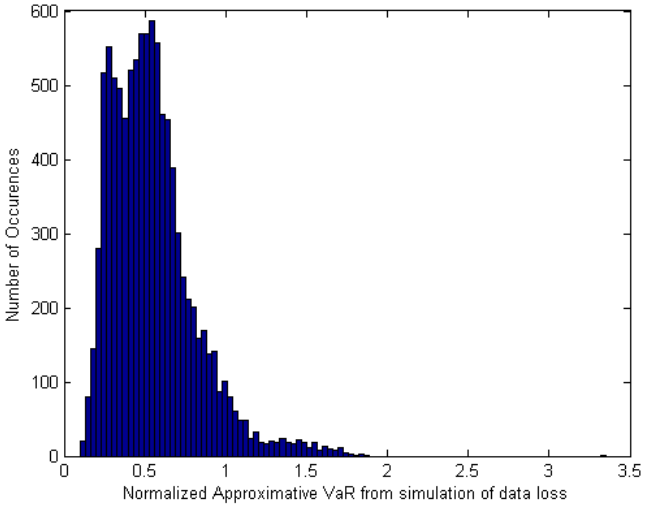


Figure 32 : *Histograms showing the distribution of the normalized approximate VaR of a cell when simulating data loss and using the MoMom-Q estimator.*

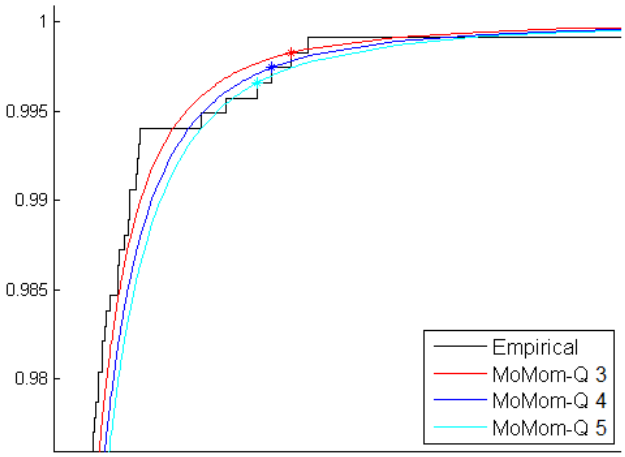


Figure 33 : *Empirical and estimated cdfs for cell 2 when using the MoMom-Q estimator with different fitted quantiles (marked with a star).*

While the MoMom-Q is able to smooth out the peaks (notice that the three peaks are still existent in Figure 32), the VaR seems to be a underestimated in general. This effect is seen for all cells, and is related to the "robustness margin" that was explained in section 4.3 (i.e. that the fitted quantile never is higher than the fifth greatest empirical loss). When the number of available losses decreases, this means that also the fitted quantile will decrease (obviously $1 - 4/N_{losses}$ decreases together with N_{losses}, and the cdf is nondecreasing). Remember that the cells with more than 1,000 ORX losses generally also have quite a few registered internal losses, and therefore also a high estimated Poisson intensity. This means that the $1 - \alpha/\mathrm{E}[N]$-quantile (see equation (1)) will be close to 1 and the fitted quantile will be x_5.

As the shape parameter is generally underestimated by the MoMom-Q estimator, since $\hat{\xi}_{MoM} < 1/2$ and $\hat{\xi}_{ML} > 1/2$ for all cells (see Table 3), the estimated GPD will tend to overestimate the quantiles below the fitted quantile and underestimate the quantiles above the fitted quantile. This behaviour is illustrated in Figure 33, where the edf for cell 2 is shown along with the fitted GPD when using the MoMom-Q estimator with a minimum fitted "empirical index" of 3, 4 and 5 respectively.

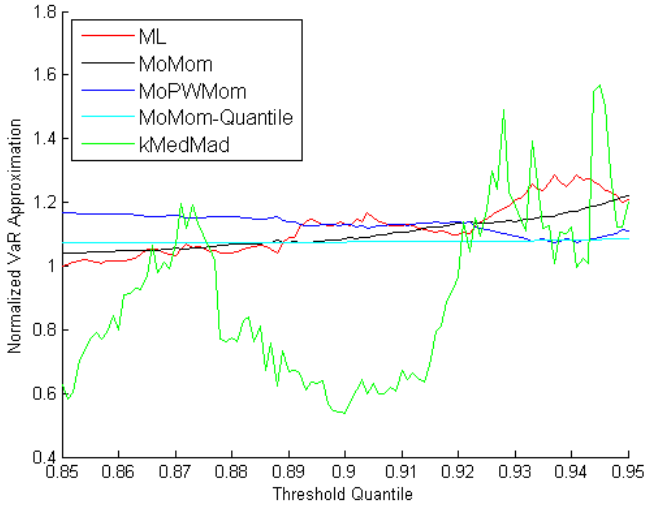


Figure 34 : *VaR-approximations of the aggregated loss with respect to the chosen threshold quantile using five different estimators.*

5.7 Threshold Stability

When calculating the relative errors from the VaR-approximations with respect to different plausible thresholds (say between the 0.85- and the 0.98-empirical quantiles), one finds the same robustness hierarchy as in the previous section. That is, the ML estimator and the kMedMad estimators are very unstable in this regard, MoMom-Q is very stable (as expected), and the MoPWMom and MoMom estimators can be placed somewhere in between. This is illustrated in Figure 34 where the VaR-approximation is plotted as a function of the threshold quantile for a given cell.

The stability of the MoMom-Q estimator with respect to simulated data loss is also very independent of the chosen threshold. Figure 35 shows the sum of the relative errors of the VaR-estimate from all cells with more than 2,000 ORX losses, when performing the same calculations as in the previous section for thresholds ranging from the 0.84 to the 0.96 empirical quantiles. While the instability of the ML estimator increases with the threshold, the MoMom-Q can be seen to be comparatively stable in this regard. Note, however, that these studies have used the approximate VaR from section 2.5, which is exactly the quantity that MoMom-Q tries to fit to the corresponding empirical estimate. As was previously discussed in section 5.3, the validity of this approximation is somewhat diminished when using MoMom-Q estimates for cells with larger Poisson intensities.

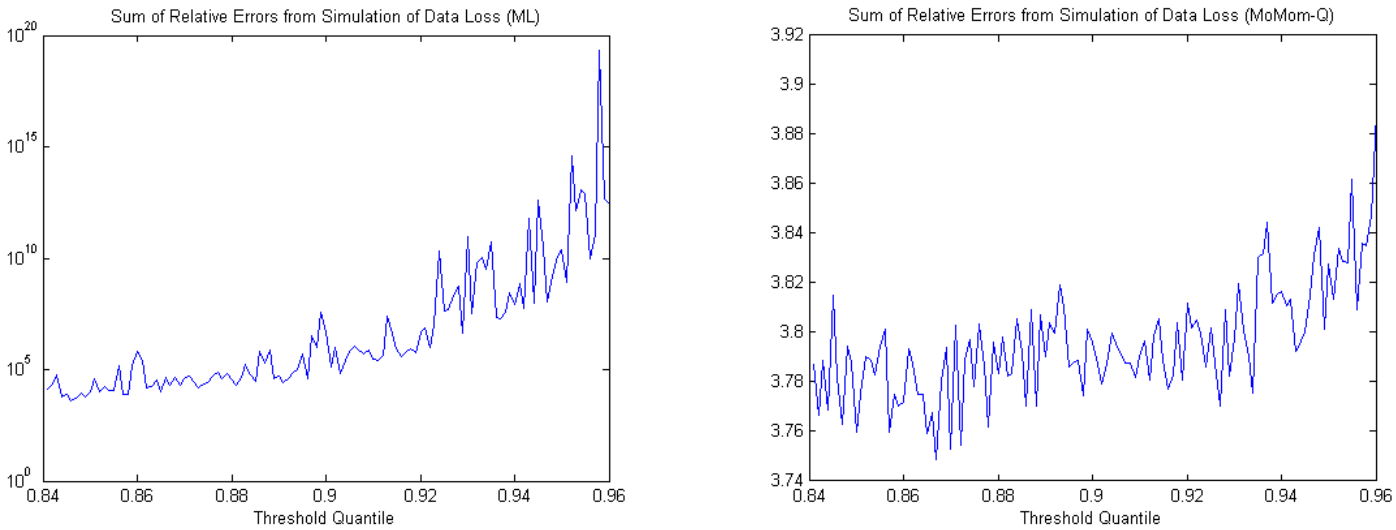


Figure 35 : *The sum of the relative errors from different cells when simulating data loss with respect to different chosen thresholds.*

5.8 Cell Aggregation

To support the incorporation of BEICFs in the model (see paragraph 676 in International Convergence of Capital Measurement and Capital Standards), the aggregation of BLs will be performed so as to match the internal organization structure. This means that only four BLs will be used: Corporate Finance (CF), Wholesale Banking (WB), Retail Banking (RB) and Wealth Management (WM). The mapping can be seen in Table 1.

The number of relevant ORX observations available for each of the now 28 cells left, is a critical factor when evaluating further aggregation. As Figure 36 illustrates, the issue of data sufficiency will mostly be focused on the BL CF.

There are two primary ways in which the aggregation will affect the total capital allocation. First, remember that the LDA assumes independence between the number of events and the severities in

Model BL	*ORX BL, Level 1*	*ORX BL, Level 2*	*Code*
Corporate Finance	Corporate Finance	Corporate Finance	BL0101
Wholesale Banking	Trading & Sales	Equites	BL0201
		Global Markets	BL0202
		Corporate Investment	BL0203
		Treasury	BL0204
	Commercial Banking	Commercial Banking	BL0301
	Clearing	Cash Clearing	BL0501
	Agency Services	Custody Services	BL0601
Retail Banking	Retail Banking	Retail Banking	BL0301
Wealth Management	Asset Management	Fund Management	BL0703
	Retail Brokerage	Retail Brokerage	BL0801
	Private Banking	Private Banking	BL0901

Table 1 : *Mapping table for BLs.*

each individual cell. On the other hand, the dependence modeling in between cells is generally quite conservative and typically goes as far as assuming perfect correlation. As a consequence, a more granular model will often go hand in hand with greater dependence, and therefore also higher capital estimates.

By contrast, let us study the effect of aggregating two cells which roughly can be characterized as low frequency/high severity and high frequency/low severity respectively. This means that the highest quantiles of the severity distribution of the aggregated cell will be estimated using observations with high severities from the first cell, while the intensity of the event variable will be driven by losses from the second cell. The risk is then that the probability of a high loss severity is overestimated, since the magnitude of the greatest loss observations from the first cell might be unlikely to ever see in practice from the second cell. Note that the number of available losses most often is less than $\mathrm{E}[N]/0.001$ (see equation (1)) and that the severity distribution have to be "extrapolated" in the most important quantiles. Therefore, not only the values of the highest losses, but also their heavy-tailedness, will be important when estimating the upper tail of the severity distribution. Of course, the final effect of aggregations of this type will be dependent on both the specific observations available and the chosen model in each cell.

Since the effect of the first aggregation aspect is mainly related to inadequate modeling as a result of data insufficiency and regulation standards, while the second aspect examplifies an often avoidable failure of the model to characterize real world risk features, it is appropriate to focus on eliminating the effects of the second aspect. In other words, one should hope that the first issue can be overcome by using a more refined dependence modeling than we will study in this thesis, while the second is due to a more direct failure of the actual cell aggregation. We therefore applied the two-sample Kolmogorov-Smirnov

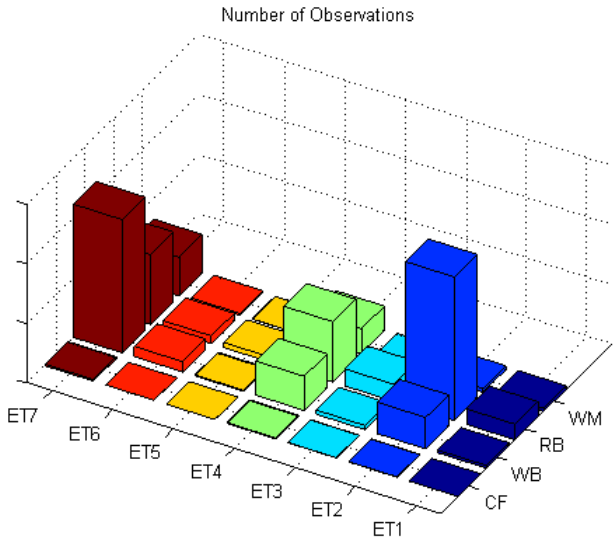


Figure 36 : *The number of ORX observations when using the BL-mapping from Table* 1 *and the original level* 1 *ETs.*

ET1 & ET2	
BL	p-value
CF	Inconclusive
WB	2.6828e−38
RB	2.4550e−213
WM	2.6248e−24

ET5 & ET6	
BL	p-value
CF	Inconclusive
WB	8.9859e−4
RB	1.8622e−8
WM	0.0545

Table 2 : *P-values given by performing the two-sample Kolmogorov-Smirnov on the loss data from two cells with different ETs and the same BL.*

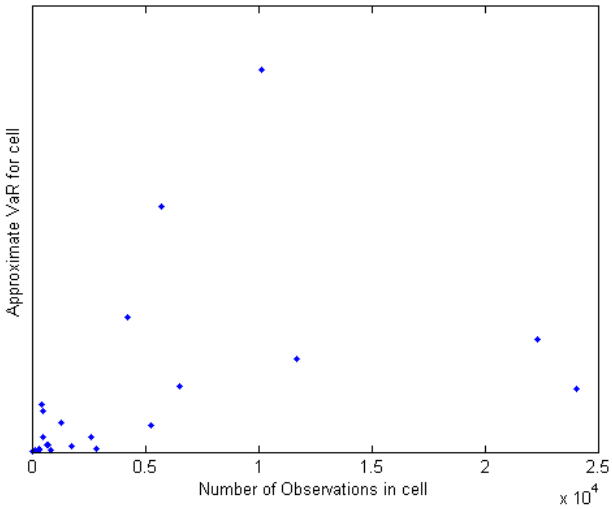


Figure 37 : *The approximate capital requirements from the individual cells in the final model as a function of the number of available ORX losses.*

test on the ORX losses from ETs which are aggregated in Aue and Kalkbrenner (2006), that is, internal and external fraud, as well as Disasters & Public Safety and Technology & Infrastructure. The results can be seen in Table 2.

Obviously, aggregation is not supported for any of the suggested pairs of cells, and the quantitative result is inconclusive for the BL CF since too few loss observations are available. Despite this, our final model will aggregate all ETs for CF, since the internal losses are not sufficiently many to allow for a more risk sensitive model. Furthermore, we will use a conservative intensity estimate for two other cells which have no relevant internal losses by simply performing the calculations as if we had one internal observation for each of these cells. Since these issues concerns cells with very few internal losses anyway, our judgement is that these modeling decisions will have a limited effect on the total regulatory capital. As is shown in Figure 37, the largest capital estimates generally come from cells with a large number of available losses.

5.9 Results on Correlation Modeling

We will in this section model the severity distribution as a piecewise function with a GPD-tail, empirical body, and a threshold at the empirical 0.9-quantile. The parameter estimates will be given by the MoMom-Q estimator, which makes the threshold choice relatively insignificant, as previously discussed in section 5.7. The choice to use the MoMom-Q estimator is further motivated by refering to its UTAD-performance (see section 5.2), its implied rate of convergence for MC estimates (see section 5.4) and its robustness properties and insensitivity towards outliers of high severity (see sections 5.5 and 5.6). The mean of the severity distribution can then be written

$$\mathrm{E}[X] \leq \frac{1}{10}\mathrm{E}[X_{GPD}] + u = \frac{1}{10}\frac{\beta}{1-\xi} + u \leq \frac{\beta}{5} + u.$$

Correspondingly, we get

$$\begin{aligned} 10\mathrm{E}[X^2] &\geq \mathrm{E}[(X_{GPD}+u)^2] = \mathrm{E}[X_{GPD}^2] + 2u\mathrm{E}[X_{GPD}] + u^2 \\ &= \frac{2\beta^2}{(1-\xi)(1-2\xi)} + 2u\frac{\beta}{1-\xi} + u^2, \end{aligned} \tag{19}$$

and hence we have

$$\eta \leq \sqrt{10}\left(\frac{\beta}{5}+u\right) \Big/ \sqrt{\frac{2\beta^2}{(1-\xi)(1-2\xi)} + 2u\frac{\beta}{1-\xi} + u^2}. \tag{20}$$

Using the cell aggregation from the previous section, we have $u < \beta < 8u$ for all but one cell, and we will thereby make use of the very crude approximation $\beta \approx u$ to arrive at

$$\eta \lesssim \frac{6}{5}\sqrt{\frac{10(1-\xi)(1-2\xi)}{5-7\xi+2\xi^2}} = J(\xi). \tag{21}$$

To increase robustness, we will model all the six severity distributions from cells with less than 500 ORX losses with the WBD. These cells give about 5% of the total regulatory capital when assuming perfect correlation. Note that this modeling choice means that the parameter estimates of a GPD-tail always are based on at least 50 observations.

By studying Figure 38 and noting that

$$\eta(X_i) = \frac{1}{\sqrt{1 + \mathrm{Var}(X_i)/\mathrm{E}[X_i]^2}} \leq 1,$$

we see that the inequality in (21) is only useful for ξ very close to 1/2. However, those cells that contribute the most to the total capital requirement are typically very heavy-tailed, and in our final model the two cells associated with about a third of the total allocation when assuming perfect correlation both had $\xi > 0.485$, which corresponds to $J(\xi) \leq J(0.485) \approx 0.3274$. By using their exact values of β and u in (20) we get $\eta \leq 0.0199$ and $\eta \leq 0.0374$ respectively, i.e. even stricter upper limits on η. Note that these numbers are in line with the previously mentioned values from Frachot et al. (2004). We further implemented a model where a Gaussian Copula was specified for the event variables associated with different cells, and where all non-diagonal elements in the correlation matrix of the Copula were given the same value. The resulting capital estimates supported the notion that the correlation between aggregated loss variables will be very small in this case, since no significant difference was registered in the total capital requirement as the correlation parameter was allowed to vary between 0 and 1.

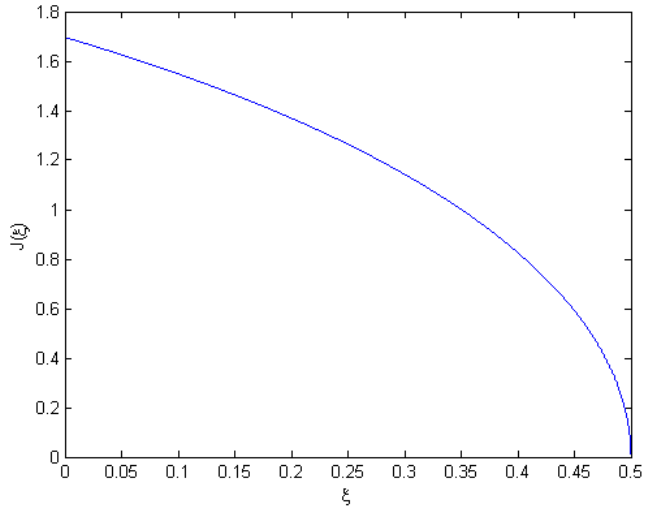


Figure 38 : *The function* $J(\xi)$ (*see equation* (21)) *limiting the factor* η *when assuming* $\beta = u$.

Of course, these heuristics does not say anything about any real-world correlation, but merely displays that in order for the model to recognize a non-negligible correlation of the aggregated losses in between cells, it will probably have to include some kind of severity dependence.

6 Conclusions

The main appeal of the data weighting method discussed in section 3.2 is that it can manage to unify the empirical cdfs of two very different sets of loss data, without scaling or discarding any individual severities. However, as has been shown, the error function proposed in Aue and Kalkbrenner (2006) is far from optimal since it disregards the importance of the high severity quantiles. Even though it would be easy to modify the error function so that it gives a larger weight to the highest severity quantiles, the procedure is complicated by the fact that the severities of the highest losses will vary considerable depending on from which cell the loss is drawn. Moreover, it is often impractical to perform separate weightings of losses from different cells due to the scarcity of data.

While section 5.1 supports previously published papers that advocates the POT method, there are other characteristics than data fit that should be taken into consideration when deciding on how to model the severity distributions. These include both computational speed and stability of the regulatory capital estimates.

As was illustrated in section 5.5, conventional threshold estimators are not suited for the particular needs of operational risk modeling. They have been designed with the intention of motivating the placement of the boundary separating the vaguely defined extreme value domain from the standard values of, in our case, the severity distribution. The aim has not been stability of high quantile estimates with respect to a continuously updated data set, and so their irregular behaviour in these applications should not come as a surprise.

The risk measure approximations presented in section 2.5 are extremely useful when one wants to receive a fast estimation of the effect that a certain modeling choice has on the capital requirements. The precision of the approximation when estimating the total capital requirement is very dependent on whether or not those cells with the highest capital requirements fall into the "poor-performing" category described in section 5.3, which ultimately depends on several modeling choices including the chosen severity distribution, parameter estimator and cell aggregation. Also remember that a straightforward application of the approximation is limited to the case when we assume perfect correlation in between cells.

The analysis in section 5.4 shows that the relative errors of the MC VaRs will be approximately linear with regard to the shape parameter of the severity distribution. Keep in mind, however, that the actual values of the standard deviations will depend on both the intensity of the Poisson distribution, and the severity distribution (including the threshold and the scale parameter when using the GPD). Furthermore, we cannot assume that the convergence in equation (17) to the normal distribution is fast enough for our purposes, and hence the estimated relative error might not give us the complete picture of the convergence characteristics.

Section 5.6 obviously speaks in favour of the estimators based on moments, since they more or less eliminates the extreme overestimates of the regulatory capital that are possible when using the ML estimator. The fact that the capital estimates will be somewhat biased when using the MoMom-Q estimator with too small a number of observations (see Figure 32), shows that the fitted empirical quantile should be dependent on N_{losses}. If the fitted quantile is chosen at approximately the same level, no matter how many observations are available, the index of the fitted observation (when ordering the observations as $x_n \leq ... \leq x_2 \leq x_1$) will increase proportionally to N_{losses} (so that the FSBP of the fitted empirical quantile is independent of N_{losses}). This is natural from a robustness perspective, since over any given time period, a larger number of new observations is expected to be added to those cells that already contain a large number of losses.

Besides its robustness properties, the use of the MoMom-Q estimator can also be motivated by its UTAD-performance (see section 5.2) and its implied rate of convergence for MC estimates (see section 5.4). However, the biggest argument in favour of the estimator is that it attempts to fit, to empirical data, exactly that severity quantile which to a fairly good approximation will determine the regulatory capital.

The decision on a suitable cell aggregation is still a very open subject, and it is not made clearer by the different specifications of ETs and BLs given by ORX and the Standardized Approach. Furthermore, these divisions will seldom perfectly reflect the business divisions of each specific bank, and some compromising aggregation will often have to be specified, perhaps supported by estimates of upper limits on capital estimates from specific cells. Unfortunately, the analysis performed in section 5.8 confirms that many aggregations often are as hard to motivate on quantitative grounds as on qualitative (notice that the aggregations which are tested and rejected in Table 2 are implemented in Aue and Kalkbrenner (2006)). For instance, internal frauds are often very different in nature from external frauds simply because the former usually is based on abuse of some authority or resource which cannot be accessed from the outside.

The problem of cell aggregation is closely knit to the problem of modeling correlation, since the better you can model the real world correlations, the less the cell aggregation will effect the capital estimates. Section 5.9 shows that simply adding dependence between the event variables will have a very small effect on the total capital estimates. Further, as was concluded in Aue and Kalkbrenner (2006), there are no indications that it would be wise to relax the standard assumptions of independence between the variables in each specific cell. Therefore, the use of a common shock model, first applied to loss models in Lindskog and McNeil (2003), seems to be a reasonable alternative in a model aiming for a non-negligible correlation effect.

7 Suggestions for Further Studies

From a regulatory point of view, it is not optimal to limit the shape parameter of the GPD-tail to only take on values $\leq 1/2$. Using, for instance, the ML estimator, will on the other hand often increase the computational time required to obtain stable MC estimates severely. This is because the cells which contributes the most to the total capital requirement typically have both large intensities and very heavy tails, which means that each simulated yearly loss will require a larger number operations, and that the number of such simulations needed for a specified accuracy will be relatively large. One possibility would be to study how importance sampling could be used in the LDA model to speed up the MC estimates. Applications of importance sampling to a model identical to the single cell LDA with a heavy-tailed severity distribution has previously been studied in Dupuis et al. (2007).

It is probable that improvements of the analytical VaR approximation in the case of large intensities and small shape parameters can be made by making adjustments of the approximation based on other asymptotics than the subexponential. The problem can be formulated as a search for an analytical inversion of the expression

$$\mathrm{P}(S > s) = \mathrm{E}[\mathrm{P}(X_1 + ... + X_N > s)] = \sum_{n=1}^{\infty} \mathrm{P}(X_1 + ... + X_n > s) f_{ev}(n),$$

where the approximations in section 2.5 uses $\mathrm{P}(X_1 + ... + X_n > s) \approx n\mathrm{P}(X > s)$. As already mentioned, this approximation fails for sufficiently large n and sufficiently small ξ, in which case it would be advisable to examine the usefulness of e.g. central limit theorem approximations.

Further ideas for numerical approximations of the aggregated loss distribution can be found in Frachot et al. (2001) which reviews both Panjer's method and the method of characteristic functions. The former results in a suggestive implicit equation for the aggregated loss distribution which can be formulated as

$$f_{agg}(s) = f_{ev}(1) f_{sev}(s) + \int_0^s \left(c_1 + c_2 \frac{x}{s}\right) f_{sev}(x) f_{agg}(s - x) dx,$$

given that

$$f_{ev}(n) = \left(c_1 + \frac{c_2}{n}\right) f_{ev}(n - 1),$$

which is satisfied for all standard event distributions.

One suggestion would be to further study how estimators similar to MoMom-Q can be constructed by using estimates of the shape parameter limited from above, combined with an estimate of the scale parameter based on matching one or several statistics from the upper quantiles of the edf and the fitted distribution. Ideally, the chosen statistics should vary together with the number of available observations as well as their characteristics to provide the desired balance between robustness and data consideration in the capital estimates.

At last, the consequences of using the edf or alternative distributions for modeling the body of the severity distributions should be studied. This is especially important when the risk measure estimates from section 2.5 fails, in which case the body of the severity distribution can be suspected to have a greater impact on the capital estimates.

8 Appendix

8.1 Abbreviations and Notation

$F(x)$ and $f(x)$ denotes a generic cdf and pdf respectively, and $\bar{F}(x) = 1 - F(x)$.

Abbreviations

BL	Business Line
(EFS)BP	(Expected Finite Sample) Breakdown Point
BEICFs	Business Environment and Internal Control Factors
cdf	Cumulative Distribution Function
CF	Corporate Finance
edf	Empirical Distribution Function
ET	Event Type
EVT	Extreme Value Theory
GEV	Generalized Extreme Value
GPD	Generalized Pareto Distribution
(E)IF	(Empirical) Influence Function
KS	Kolmogorov-Smirnov
LDA	Loss Distribution Approach
LND	LogNormal Distribution
MC	Monte Carlo
ML	Maximum Likelihood
MoMom(-Q)	Method of Moments(-Quantile)
MoPWMom/PWM	Method of Probability Weighted Moments/Probability Weighted Moments
pdf	Probability Distribution Function
POT	Peaks Over Threshold
RB	Retail Banking
SD	Standard Deviation
SF	Sensitivity Function
UTAD	Upper Tail Anderson-Darling
WB	Wholesale Banking
WBD	Weibull Distribution
WM	Wealth Management

8.2 Approximations of Risk Measures

First note that the GPD is subexponential and that equation (2), given the Poisson distribution with intensity λ, evaluates to

$$\sum_{n=0}^{\infty}(1+\epsilon)^n\frac{\lambda^n e^{-\lambda}}{n!}=\sum_{n=0}^{\infty}\frac{((1+\epsilon)\lambda)^n e^{-\lambda}}{n!}=e^{\lambda\epsilon}\sum_{n=0}^{\infty}\frac{((1+\epsilon)\lambda)^n e^{-\lambda(1+\epsilon)}}{n!}=e^{\lambda\epsilon},$$

which is finite for all $\epsilon>0$. Theorem 1.3.9 in Embrechts et al. (1997) then directly gives

$$F_{agg}(s)\to 1-\mathrm{E}[N](1-F_{sev}(s)),\text{ as } s\to\infty,$$

and further

$$\mathrm{VaR}_\alpha(S)=F_{agg}^{-1}(1-\alpha)\approx F_{sev}^{-1}\left(1-\frac{\alpha}{\mathrm{E}[N]}\right).$$

Modeling $F_{sev}(x)$ as the edf with a GPD tail now gives

$$F_{sev}^{-1}(x)=u+G_{\xi,\beta}^{-1}\left(1-(1-x)\frac{N_{losses}}{N_{losses>u}}\right),$$

whenever $x>1-N_{losses>u}/N_{losses}$, and so,

$$\mathrm{VaR}_\alpha(S)\approx u+G_{\xi,\beta}^{-1}\left(1-\frac{N_{losses}}{N_{losses>u}}\frac{\alpha}{\mathrm{E}[N]}\right)=u+\frac{\beta}{\xi}\left(\left(\frac{N_{losses>u}}{N_{losses}}\frac{\mathrm{E}[N]}{\alpha}\right)^{\xi}-1\right).$$

The estimate of $\mathrm{ES}_\alpha(S)$ can now simply be obtained by using this VaR-estimate in the definition of $\mathrm{ES}_\alpha(S)$.

8.3 The Mean of the Generalized Pareto Distribution

Provided that $0<\xi<1$, we have

$$\begin{aligned}\mu&=\int_0^\infty x g_{\xi,\beta}(x)dx=\frac{1}{\beta}\int_0^\infty x\Big(1+\xi\frac{x}{\beta}\Big)^{-\frac{1}{\xi}-1}dx\\&=-\left[x\Big(1+\xi\frac{x}{\beta}\Big)^{-\frac{1}{\xi}}\right]_0^\infty+\int_0^\infty\Big(1+\xi\frac{x}{\beta}\Big)^{-\frac{1}{\xi}}dx\\&=\left[\frac{\beta}{\xi-1}\Big(1+\xi\frac{x}{\beta}\Big)^{1-\frac{1}{\xi}}\right]_0^\infty=\frac{\beta}{1-\xi}.\end{aligned}$$

The calculations are almost identical in the case of $\xi<0$, except that the upper limit of the integrals is $-\beta/\xi$.

8.4 The Variance of the Generalized Pareto Distribution

Assuming $0<\xi<1/2$, we have

$$\sigma^2+\mu^2=\int_0^\infty x^2 g_{\xi,\beta}(x)dx=\frac{1}{\beta}\int_0^\infty x^2\Big(1+\xi\frac{x}{\beta}\Big)^{-\frac{1}{\xi}-1}dx=-\left[x^2\Big(1+\xi\frac{x}{\beta}\Big)^{-\frac{1}{\xi}}\right]_0^\infty+\int_0^\infty 2x\Big(1+\xi\frac{x}{\beta}\Big)^{-\frac{1}{\xi}}dx$$
$$=-\left[2x\frac{\beta}{1-\xi}\Big(1+\xi\frac{x}{\beta}\Big)^{1-\frac{1}{\xi}}\right]_0^\infty+\int_0^\infty\frac{2\beta}{1-\xi}\Big(1+\xi\frac{x}{\beta}\Big)^{1-\frac{1}{\xi}}dx=-\left[\frac{2\beta^2}{(1-\xi)(1-2\xi)}\Big(1+\xi\frac{x}{\beta}\Big)^{2-\frac{1}{\xi}}\right]_0^\infty$$
$$=\frac{2\beta^2}{(1-\xi)(1-2\xi)}.$$

This immediately gives the sought after expression for σ^2. The calculations are almost identical in the case of $\xi<0$, except that the upper limit of the integrals are $-\beta/\xi$.

8.5 The $(1,0,1)$ Probability Weighted Moment of the Generalized Pareto Distribution

Assuming $0<\xi<2$ gives

$$M_{1,0,1}=\int_0^\infty x g_{\xi,\beta}(x)(1-G_{\xi,\beta}(x))dx=\frac{1}{\beta}\int_0^\infty x\Big(1+\xi\frac{x}{\beta}\Big)^{-\frac{2}{\xi}-1}dx$$
$$=-\left[\frac{1}{2}x\Big(1+\xi\frac{x}{\beta}\Big)^{-\frac{2}{\xi}}\right]_0^\infty+\frac{1}{2}\int_0^\infty\Big(1+\xi\frac{x}{\beta}\Big)^{-\frac{2}{\xi}}dx$$
$$=\left[\frac{\beta}{2}\frac{(1+\xi\frac{x}{\beta})^{1-\frac{2}{\xi}}}{(\xi-2)}\right]_0^\infty=\frac{\beta}{2(2-\xi)}.$$

The calculations are almost identical in the case of $\xi<0$, except that the upper limit of the integrals are $-\beta/\xi$.

8.6 The Mean Excess Function

Assuming the GPD with $0<\xi<1$ we obtain

$$e(u)\bar{G}_{\xi,\beta}(u)=\int_u^\infty(x-u)g_{\xi,\beta}(x)dx=\frac{1}{\beta}\int_u^\infty x\Big(1+\xi\frac{x}{\beta}\Big)^{-\frac{1}{\xi}-1}dx-u\bar{G}_{\xi,\beta}(u),$$

where

$$\frac{1}{\beta}\int_u^\infty x\Big(1+\xi\frac{x}{\beta}\Big)^{-\frac{1}{\xi}-1}dx=-\left[x\Big(1+\xi\frac{x}{\beta}\Big)^{-\frac{1}{\xi}}\right]_u^\infty+\int_u^\infty\Big(1+\xi\frac{x}{\beta}\Big)^{-\frac{1}{\xi}}dx$$
$$=u\Big(1+\xi\frac{u}{\beta}\Big)^{-\frac{1}{\xi}}+\left[\frac{\beta}{\xi-1}\Big(1+\xi\frac{x}{\beta}\Big)^{1-\frac{1}{\xi}}\right]_u^\infty$$

$$= u\bar{G}_{\xi,\beta}(u) + \frac{\beta}{1-\xi}\Big(1+\xi\frac{u}{\beta}\Big)^{1-\frac{1}{\xi}} = \bar{G}_{\xi,\beta}(u)\bigg(u + \frac{\beta}{1-\xi}\Big(1+\xi\frac{u}{\beta}\Big)\bigg),$$

which immediately gives

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

The calculations are almost identical in the case of $\xi < 0$, except that the upper limit of the integrals are $-\beta/\xi$.

8.7 The Median Excess Function

Assuming the GPD, the median excess function can be written

$$\begin{aligned} F^{(u)}(x) &= \mathrm{P}(X-u \le x | X > u) = \mathrm{P}(X \le x+u | X > u) \\ &= \frac{P(u < X \le x+u)}{P(u<X)} = \frac{G_{\xi,\beta}(x+u) - G_{\xi,\beta}(u)}{1 - G_{\xi,\beta}(u)} \\ &= \frac{(1+\xi\frac{u}{\beta})^{-\frac{1}{\xi}} - (1+\xi\frac{x+u}{\beta})^{-\frac{1}{\xi}}}{(1+\xi\frac{u}{\beta})^{-\frac{1}{\xi}}} = 1 - \left(\frac{1+\xi\frac{x+u}{\beta}}{1+\xi\frac{u}{\beta}}\right)^{-\frac{1}{\xi}} \end{aligned}$$

Inverting $F^{(u)}(x) = 1/2$ now gives equation (7).

8.8 Maximum Likelihood Estimates for the Lognormal Distribution

With $f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}$ and $Y = \ln X$, $f_Y(y)|dy| = f_X(x)|dx|$ gives us

$$f_X(x) = \frac{|dy|}{|dx|} f_Y(\ln x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2}.$$

This now gives the log likelihood function

$$\ln l(\mu,\sigma) = -\sum_{i=1}^{n} \ln x_i - \frac{n}{2}\ln 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(\ln x_i - \mu)^2,$$

from which the estimates in equations (8) and (9) follows trivially by evaluating $\frac{\partial \ln l(\mu,\sigma)}{\partial \mu} = 0$ and $\frac{\partial \ln l(\mu,\sigma)}{\partial \sigma^2} = 0$.

8.9 Maximum Likelihood Estimates for the Weibull Distribution

With $f(x) = \frac{k}{\lambda} x^{k-1} e^{-\frac{x^k}{\lambda}}$, we get the log likelihood function

$$\ln l(k,\lambda) = n\ln\frac{k}{\lambda} + (k-1)\sum_{i=1}^{n}\ln x_i - \frac{1}{\lambda}\sum_{i=1}^{n} x_i^k,$$

which further gives

$$\frac{\partial \ln l(k,\lambda)}{\partial k} = \frac{n}{k} + \sum_{i=1}^{n}\ln x_i - \frac{1}{\lambda}\sum_{i=1}^{n} x_i^k \ln x_i$$

and

$$\frac{\partial \ln l(k,\lambda)}{\partial \lambda} = -\frac{n}{\lambda} + \frac{1}{\lambda^2}\sum_{i=1}^{n} x_i^k.$$

This directly gives (11), while (10) follows by eliminating λ from $\frac{\partial \ln l(k,\lambda)}{\partial k} = \frac{\partial \ln l(k,\lambda)}{\partial \lambda} = 0$.

8.10 The Upper Tail Anderson-Darling Statistic

Integrating directly over $F(x) = z$ gives us

$$\begin{aligned}
\frac{AD}{n} &= \int_{-\infty}^{\infty} \frac{(\hat{F}(x) - F(x))^2}{(1-F(x))^2} dF(x) = \int_0^1 \frac{(\hat{F}(x) - z)^2}{(1-z)^2} dz \\
&= \int_0^{F(x_n)} \frac{z^2}{(1-z)^2} dz + \sum_{i=1}^{n-1} \int_{F(x_{n+1-i})}^{F(x_{n-i})} \frac{(i/n - z)^2}{(1-z)^2} dz + \int_{F(x_1)}^{1} dz \\
&= \Big[z + \frac{1}{1-z} + 2\ln(1-z)\Big]_0^{F(x_n)} \\
&\quad + \sum_{i=1}^{n-1} \Big[z + \frac{(i/n-1)^2}{1-z} + 2(1-i/n)\ln(1-z)\Big]_{F(x_{n+1-i})}^{F(x_{n-i})} + 1 - F(x_1).
\end{aligned}$$

The first part of the integral evaluates to

$$\Big[z + \frac{1}{1-z} + 2\ln(1-z)\Big]_0^{F(x_n)} = F(x_n) + \frac{1}{1-F(x_n)} + 2\ln(1-F(x_n)) - 1,$$

while the second is equal to

$$\sum_{i=1}^{n-1} \Big[z + \frac{(i/n-1)^2}{1-z} + 2(1-i/n)\ln(1-z)\Big]_{F(x_{n+1-i})}^{F(x_{n-i})}$$

$$= F(x_1) - F(x_n) + \frac{1}{n^2}\sum_{i=1}^{n}(1+2(n-i))\frac{1}{1-F(x_{n+1-i})} - \frac{1}{1-F(x_n)} + \frac{2}{n}\sum_{i=1}^{n}\ln(1-F(x_{n+1-i})) - 2\ln(1-F(x_n)).$$

Adding all the terms finally gives equation (14).

8.11 Correlation between Aggregated Loss Distributions

First, remember that

$$\mathrm{Corr}(S_1, S_2) = \frac{\mathrm{Cov}(S_1, S_2)}{\sqrt{\mathrm{Var}(S_1)}\sqrt{\mathrm{Var}(S_2)}},$$

while $\mathrm{E}[N_1] = \lambda_1$ and $E[N_1^2] = \lambda_1(\lambda_1 + 1)$. We then have

$$\mathrm{E}[S_1] = \mathrm{E}\left[\sum_{n=1}^{N_1} X_1^{(n)}\right] = \mathrm{E}\left[X_1^{(1)} + ... + X_1^{(N_1)}\right]$$

$$= \sum_{n=1}^{\infty} \mathrm{P}(N_1 = n)\mathrm{E}\left[X_1^{(1)} + ... + X_1^{(n)}\right]$$
$$= \mathrm{E}[X_1] \sum_{n=1}^{\infty} \mathrm{P}(N_1 = n)n$$
$$= \mathrm{E}[X_1]\mathrm{E}[N_1], \tag{22}$$

and

$$\mathrm{E}[S_1 S_2] = \mathrm{E}\left[\sum_{n=1}^{N_1} X_1^{(n)} \sum_{m=1}^{N_2} X_2^{(m)}\right] = \mathrm{E}\left[\left(X_1^{(1)} + ... + X_1^{(N_1)}\right)\left(X_2^{(1)} + ... + X_2^{(N_2)}\right)\right]$$
$$= \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \mathrm{P}(N_1 = n, N_2 = m)\mathrm{E}\left[\left(X_1^{(1)} + ... + X_1^{(n)}\right)\left(X_2^{(1)} + ... + X_2^{(m)}\right)\right]$$
$$= \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \mathrm{P}(N_1 = n, N_2 = m)\mathrm{E}\left[X_1^{(1)} + ... + X_1^{(n)}\right]\mathrm{E}\left[X_2^{(1)} + ... + X_2^{(m)}\right]$$
$$= \mathrm{E}[X_1]\mathrm{E}[X_2] \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \mathrm{P}(N_1 = n, N_2 = m)nm$$
$$= \mathrm{E}[X_1]\mathrm{E}[X_2]\mathrm{E}[N_1 N_2], \tag{23}$$

which gives

$$\mathrm{Cov}(S_1 S_2) = \mathrm{E}[S_1 S_2] - \mathrm{E}[S_1]\mathrm{E}[S_2]$$
$$= \mathrm{E}[X_1]\mathrm{E}[X_2](\mathrm{E}[N_1 N_2] - \mathrm{E}[N_1]\mathrm{E}[N_2])$$
$$= \mathrm{E}[X_1]\mathrm{E}[X_2]\mathrm{Cov}(N_1, N_2). \tag{24}$$

Furthermore, we have

$$\mathrm{E}\left[S_1^2\right] = \mathrm{E}\left[\sum_{n=1}^{N_1} X_1^{(n)} \sum_{n=1}^{N_1} X_1^{(n)}\right] = \mathrm{E}\left[\left(X_1^{(1)} + ... + X_1^{(N_1)}\right)^2\right]$$
$$= \sum_{n=1}^{\infty} \mathrm{P}(N_1 = n)\mathrm{E}\left[\left(X_1^{(1)} + ... + X_1^{(n)}\right)^2\right]$$
$$= \sum_{n=1}^{\infty} \mathrm{P}(N_1 = n)\left(\sum_{i=1}^{n} \mathrm{E}\left[\left(X_1^{(i)}\right)^2\right] + \sum_{i\neq j} \mathrm{E}\left[X_1^{(i)} X_1^{(j)}\right]\right)$$
$$= \sum_{n=1}^{\infty} \mathrm{P}(N_1 = n)\left(n\mathrm{E}\left[X_1^2\right] + n(n-1)(\mathrm{E}[X_1])^2\right)$$
$$= \left(\mathrm{E}\left[X_1^2\right] - (\mathrm{E}[X_1])^2\right) \sum_{n=1}^{\infty} \mathrm{P}(N_1 = n)n + (\mathrm{E}[X_1])^2 \sum_{n=1}^{\infty} \mathrm{P}(N_1 = n)n^2$$
$$= \left(\mathrm{E}\left[X_1^2\right] - \mathrm{E}[X_1]^2\right)\mathrm{E}[N_1] + \mathrm{E}[X_1]^2\mathrm{E}\left[N_1^2\right]$$
$$= \left(\mathrm{E}\left[X_1^2\right] - \mathrm{E}[X_1]^2\right)\lambda_1 + \mathrm{E}[X_1]^2\lambda_1(\lambda_1 + 1)$$
$$= \left(\mathrm{E}\left[X_1^2\right] + \mathrm{E}[X_1]^2\lambda_1\right)\lambda_1, \tag{25}$$

and so, by using (22) and (25)

$$\mathrm{Var}(S_1) = \mathrm{E}\left[S_1^2\right] - \mathrm{E}[S_1]^2 = \left(\mathrm{E}\left[X_1^2\right] + \mathrm{E}[X_1]^2\lambda_1\right)\lambda_1 - (\mathrm{E}[X_1]\lambda_1)^2 = \mathrm{E}\left[X_1^2\right]\lambda_1. \tag{26}$$

Finally, putting (24) and (26) together, we have

$$\mathrm{Corr}(S_1, S_2) = \frac{\mathrm{E}[X_1]}{\sqrt{\mathrm{E}\left[X_1^2\right]}} \frac{\mathrm{E}[X_2]}{\sqrt{\mathrm{E}\left[X_2^2\right]}} \mathrm{Corr}(N_1, N_2).$$

8.12 Results from Simulation of Data Loss

$\hat{\xi}_{ML}$	Rel $\mathrm{Err}_{\mathrm{ML}}$	Rel $\mathrm{Err}_{\mathrm{MoM}}$	Rel $\mathrm{Err}_{\mathrm{PWM}}$	Rel $\mathrm{Err}_{\mathrm{kMedMad}}$
0.84293	5.25949e5	8.26331	6.32106	7.67058e16
0.77781	1.593184e3	6.65541	4.98080	4.65852e10
1.04840	1.296031e3	2.30788	2.21448	2.22057e6
0.70716	1.82141e6	49.2303	21.2630	5.81353e12
1.33611	2.11541	2.89838	2.23935	1.47729e5
0.83735	5.06833e5	21.7405	10.2410	2.50962e11
1.09097	13.9247	3.22315	3.25391	5.47751e7
0.71076	4.81457e3	2.66878	3.81886	3.27132e9
0.80748	1.88124e3	4.33656	4.31866	2.82589e9
0.75457	2.98543e3	5.50400	6.23763	1.71985e10
0.97894	0.65111	1.89567	1.62856	1.37752e3
1.31250	5.26489	2.46344	2.25372	2.09841e7
0.75235	20.4394	4.10507	2.65844	6.340529e6

Table 3 : *Relative errors given by data loss simulation. All relative errors have been divided by* Rel $\mathrm{Err}_{\mathrm{MoM-Q}}$.

References

[1] International Convergence of Capital Measurement and Capital Standards, Basel Committee on Banking Supervision, www.bis.org (2006).

[2] Operational Risk - Supervisory Guidelines for the Advanced Measurement Approaches, Basel Committee on Banking Supervision, www.bis.org (2011).

[3] Operational Risk Reporting Standards, prepared by Mark Laycock, www.orx.org (2011).

[4] Aue, F., Kalkbrenner, M.: LDA at Work: Deutsche Bank's Approach to Quantifying Operational Risk, Journal of Operational Risk 1(4), 49-93 (2006).

[5] Beirlant, J., Goegebeur, Y., Teugels, J., Segers, J.: Statistics of Extremes, Wiley (2004).

[6] Böcker, K., Klüpperberg, C.: Operational VAR: A Closed Form Approximation, *RISK*, December ed., 90-93 (2005).

[7] Chernobai, A., Svetlozar, R., Fabozzi, F.: Composite Goodness-of-Fit Tests for Left-Truncated Loss Samples, Tech. rep., University of California Santa Barbara (2005).

[8] Croux, C.: Limit Behaviour of the Empirical Influence Function of the Median, Statistics & Probability Letters 37(4), 331-340 (1998).

[9] Deidda, R., Puliga, M.: Performances of Some Parameter Estimators of the Generalized Pareto Distribution over Rounded-Off Samples, Physics and Chemistry of the Earth 34(10-12), 626-634 (2009).

[10] Dupuis, P., Leder, K., Wang, H.: Importance Sampling for Sums of Random Variables with Regularly Varying Tails, ACM Transactions on Modeling and Computer Simulation 17(3), 1-21 (2007).

[11] Embrechts, P., Klüpperberg, C., Mikosch, T.: Modelling Extremal Events for Insurance and Finance, Springer-Verlag (1997).

[12] Embrechts, P., Frey, R., McNeil, A.J.: Quantitative Risk Management: Concepts, Techniques, and Tools, Princeton University Press (2005).

[13] Embrechts, P., Chavez-Demoulin, V., Neslehová, J.: Quantative Models for Operational Risk: Extremes, Dependence and Aggregation, Journal of Banking and Finance 30(10), 2635-2658 (2006).

[14] Enger, J., Grandell, J.: Markovprocesser och Köteori, KTH (The Royal Institute of Technology) (2006).

[15] Frachot, A., Roncalli, T., Georges, A.: Loss Distribution Approach for Operational Risk, Group de Recherhe Opérationnelle Crédit Lyonnais, France (2001).

[16] Frachot, A., Roncalli, T., Salomon, E.: The Correlation Problem in Operational Risk, MPRA Paper No. 38052, http://mpra.ub.uni-muenchen.de/38052/ (2004).

[17] Galfond, G.J.: Robust Estimation of Extreme Quantiles, Technical Report No.60, Department of Statistics, Stanford University, (1982).

[18] Graham J.: Cultural Aspects of Operational Risk, OpRisk & Compliance (2008).

[19] Gut, A.: An Intermediate Course in Probability, Springer (2009).

[20] Hill, B.M.: A Simple General Approach to Inference About the Tail of a Distribution, The Annals of Statistics 3(5), 1163-1174 (1975).

[21] Hosking, J.R.M., Wallis, J.R.: Parameter and Quantile Estimation for the Generalized Pareto Distribution, Technometrics 29(3), 339-349 (1987).

[22] Huisman, R., Koedijk, K.G., Kool, C.J.M., Palm, F.: Tail-Index Estimates in Small Samples, Journal of Business & Economic Statistics 19(1), 208-216 (2001).

[23] Hult, H., Svensson, J.: Efficient Calculation of Risk Measures by Importance Sampling - the Heavy-Tailed Case, preprint at arXiv:math/0909.3335v1 [math:PR] (2009).

[24] Lindskog, F., McNeil, A.J.: Common Poisson Shock Models: Applications to Insurance and Credit Risk Modelling, ETH Zurich, ASTIN Bulletin 33(2): 209-238 (2003).

[25] Marazzi, A., Ruffieux, C.: The Truncated Mean of an Asymmetric Distribution, Computational Statistics & Data Analysis 32(1), 79-100 (1999).

[26] Nasser, M., Alam, M.: Estimators of Influence Functions, Communications in Statistics - Theory and Methods 35(1), 21-32 (2006).

[27] Obdyke, J., Cavallo, A.: Estimating Operational Risk Capital: the Challenges of Truncation, the Hazards of MLE, and the Promise of Robust Statistics, Journal of Operational Risk 7(3), 3-90 (2012).

[28] Peng, L., Welsh, A.H.: Robust Estimation of the Generalized Pareto Distribution, Extremes 4(1), 53-65 (2001).

[29] Pickands, J.: Statistical Inference using Extreme Order Statistics, The Annals of Statistics 3(1), 119-131 (1975).

[30] Quenouille, M.H.: Notes on Bias in Estimation, Biometrika 43 (3/4), 353-360 (1956).

[31] Riess, R.D., Thomas, M.: Statistical Analysis of Extreme Values 3rd ed., Birkhauser (2007)..

[32] Ruckdeschel, P., Horbenko, N.: Robustness Properties of Estimators in Generalized Pareto Models, Technical Report No. 182, Fraunhofer ITWM, Kaiserslautern (2010).

[33] Ruckdeschel, P., Horbenko, N.: Yet Another Breakdown Point Notion: EFSBP, Metrika 75(8), 1025-1047 (2012).

[34] Scarrott, C., MacDonald, A.: A Review of Extreme Value Threshold Estimation and Uncertainty Quantification, Revstat - Statistical Journal 10(1), 33-60 (2012).

[35] Tursunalieva, A., Silvapulle, P.: Estimation of Operational Risks using a Semi-Parametric Approach, Department of Econometrics and Business Statistics, Monash University (2011).

TRITA-MAT-E 2013:58
ISRN-KTH/MAT/E—13/58-SE

www.kth.se