

A Comparison of Loss Aggregation Methods for Operational Risk

Grigory Temnov^{† ‡}, Richard Warnung^{† ‡ §}

Abstract

In this paper we address the problem of modeling and measuring operational risk. First, we introduce general ideas about operational risk — a brief description of the methodology for selecting models for loss severity and frequency is given. Then we turn to the problem of loss aggregation. Coming from insurance mathematics, there are in general three popular and commonly used ways to aggregate operational risk: a Monte Carlo approach, an approach based on the Fourier transformation, and a recursion approach. These three methods and their application are therefore discussed in this paper. For the recursion approach a reinterpretation of the portfolio credit risk model Credit Risk⁺ is used. The three methods are compared concerning speed and accuracy.

Keywords: Operational risk, Loss models, Risk aggregation, VaR, Recursive methods, FFT.

Acknowledgment: The authors thank two anonymous referees for advice that helped to significantly improve this article.

1 Operational risk data: models for severity and frequency

We start with a short overview of the methods for modeling severity and frequency of operational losses and outline the common methodological background before passing to the loss aggregation.

1.1 Basic assumptions on the data

Usually, operational risk (*OpRisk*) data keep their common traits in whichever financial institution they are observed — heavy-tailedness of loss distributions, the phenomenon of rare but very severe losses, especially for some particular lines of business, appears in any OpRisk database. At the same time, some peculiarities of business processes, responsible for loss occurrences, are present in every financial structure, and may significantly influence loss distributions for certain business lines and change the final VaR estimation dramatically. In our investigation we rely on two historical databases of operational losses: the *internal*

[†]Vienna University of Technology, Institute for Mathematical Methods in Economics, Financial and Actuarial Mathematics, Wiedner Hauptstraße 8-10/105-1, A-1040 Vienna, Austria

[‡]This work has been done under the financial support by Christian Doppler Laboratory for Portfolio Risk Management (PRisMa Lab) <http://www.prismalab.at/>

[§]Financial support by WWTF MA13 is gratefully acknowledged

data of our selected financial institution (in the following *the Bank*), and the *external data*. The latter contains the loss data from a group of banks and could be an important contribution, especially for the analysis of losses that have not yet occurred to the Bank. The task of mixing internal and external data properly poses a very important and non-trivial problem. The use of external data becomes essential when there are not sufficient internal loss records to produce satisfactory distribution fits. At the same time, external data must not hide the peculiarities of the internal data.

The problem of choosing optimal probability distributions for fitting observed heavy-tailed loss data is the subject of extreme value theory (EVT). The problem of keeping the balance between internal and external data, while taking them both into account, relates to *credibility theory*. In this section we refer to results of EVT, and also to the foundations of the *bayesian methodology*, which is one of the key points of credibility theory.

Let us make some remarks about the structure of the data. The loss data is classified both by lines of business (*BL's*) and types of loss events (*ET's*), so that each loss value corresponds to one particular BL and a particular ET. The division by BL's is more important for the problem of the calculation of regulatory capital for OpRisk, as, according to Basel II recommendations (Basel Committee of Banking Supervision 2001), it is the quantile of the aggregate loss distribution for each separate line of business that should be analyzed at first. But the division by ET's can also be taken into account dealing with problems such as insurance; i.e. if the Bank wishes to insure the losses corresponding to certain BL's and ET's.

An important part of the data mixing is the bayesian inference: if internal data is insufficient, then the external data is used to estimate prior distributions of parameters for a severity distribution and, subsequently, bayesian estimators using internal data are applied to obtain posterior distributions of the corresponding parameters.

1.2 Estimating parameters for loss severity and frequency

Applying extreme value theory to severity distributions: Extreme value theory, originated in the 1920's in the works of Fisher and Tippett (Fisher & Tippett 1928), plays an important role in actuarial problems. Two distribution families play a major role in EVT — the generalized extreme value (GEV) and the generalized Pareto distribution (GPD). The GEV describes the limit distribution of normalized maxima, while the GPD appears as the limit distribution of scaled excesses over a high threshold. Dealing with GPD one should keep in mind that it is an ideal model for large infrequent losses displaying a better fit (a) the larger the sample size is, (b) the higher the threshold is, and (c) the closer the mechanism of loss generation is to the unbiased ideal process (e.g., concerning the possibility of extremely high losses far beyond the total value of all assets of the Bank). Clearly, in practice one can never expect the above requirements to hold exactly and simultaneously. Hence, the real distribution of large losses can deviate from GPD significantly and another theoretical distribution can display a better fit. That's why considering each

business line we always search for an alternative for the GPD; in the present investigation the three-parameter Weibull distribution is used as an alternative to GPD.

The GPD is parametrized such that the distribution function is given by

$$G_{\xi,\mu,\beta}(x) = 1 - \left(1 + \xi \frac{x - \mu}{\beta}\right)^{-1/\xi}, \quad \text{for } x > \mu, \quad (1.1)$$

where $\mu \geq 0$ is the location parameter (in our approach it coincides with the selected threshold), $\beta > 0$ is the scale parameter and $\xi > 0$ is the shape parameter.

The distribution function of the Weibull distribution, in its generalized form — with shape parameter $\xi > 0$, scale parameter $\beta > 0$ and location parameter $\mu \geq 0$ — can be written as

$$W_{\xi,\mu,\beta}(x) = 1 - e^{-((x-\mu)/\beta)^\xi}, \quad \text{for } x > \mu. \quad (1.2)$$

In order to obtain severity distributions we use the peaks-over-threshold (*POT*) method, see e.g. (Embrechts, et al. 1997). Standard diagnostics tools for heavy tailed distributions, such as ME-plots and QQ-plots, are used for the preliminary analysis of the data and the initial intuitive choice of the severity distribution.

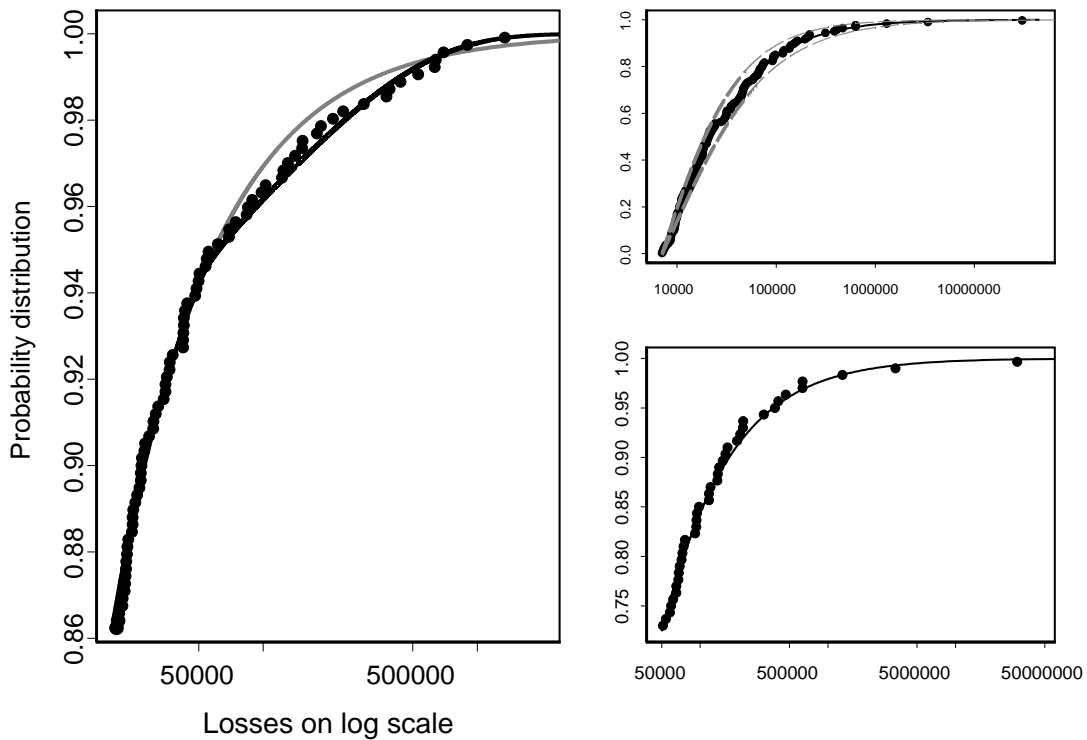
Furthermore, we find it useful to apply additional graphical diagnostics in order to figure out how precisely the data follows the GPD law with estimated parameters, and also to trace possible evolutions of the data within the observed time period. Namely under the hypothesis, that the size of losses exceeding some threshold u follows GPD, residuals in the sense of (Cox & Snell 1968), i. e.

$$W_i := \frac{1}{\xi} \log \left(1 - \frac{L_i^u - \mu}{\beta - \xi(u - \mu)}\right),$$

where L_i^u denotes the i th loss exceeding u , should be i.i.d. unit exponentially distributed. This could be checked using graphical diagnostics, if we create a QQ-plot of the ordered residuals W_i against exponential quantiles. Additionally, a plot of the residuals W_i in chronological order, superimposed with a smooth curve fit (which is based on robust locally linear fits), see Figure 2, indicates whether some evolution of the residuals W_i in time exists (here we follow the methodology proposed in (McNeil & Saladin 2000)). We notice that for certain business lines GPD overestimates the heavy-tailedness of the real loss distribution. One of the plots in Figure 1 illustrates this effect. Namely, in the case of BL 2 (left plot) we see a better fit of the Weibull distribution (confirmed by a Kolmogorov–Smirnov test), which with certain parameters has lighter tails than the GPD.

The tendency of the GPD to overestimate the probability of extremely large losses can intuitively be understood. Obviously, the GPD extrapolates the loss distribution to the area beyond the largest historical loss and of course does not take into account the fact that extremely large losses (e.g., larger than the sum of all assets) just cannot occur to the Bank.

Figure 1: Top right: BL 1 - GPD fit with confidence intervals, bottom right — fitted tail of BL 1, Left: BL2 — Comparison of GPD fit(grey) with Weibull fit (black) for the tail.



Loss Occurrences: To model counting processes, homogeneous and inhomogeneous Poisson processes, as well as negative binomial distributions arising from mixing the intensity of a Poisson process with a gamma distribution were considered in our analysis.

Remark 1.1. Working with the negative binomial distribution we use the following parametrization:

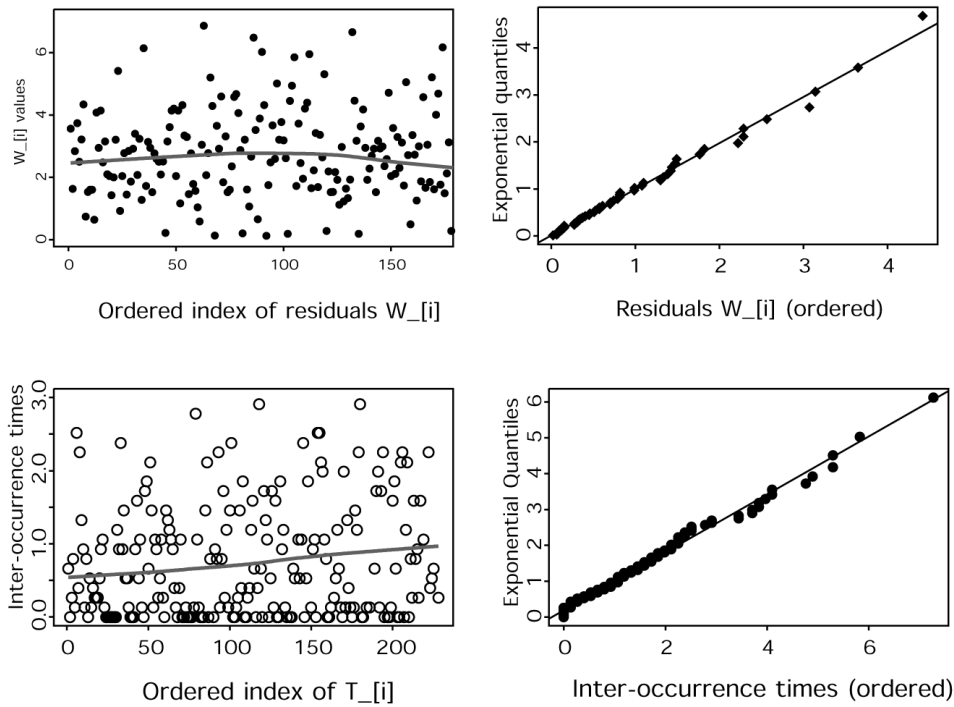
$$\mathbf{P}(N = k) = \binom{k+r-1}{k} \left(\frac{1}{1+b}\right)^r \left(\frac{b}{1+b}\right)^k, \quad (1.3)$$

with $r > 0$, $b > 0$. For the expectation and the variance we have $\mathbf{E}(N) = rb$ and $\text{Var}(N) = rb(b+1)$.

The graphical diagnostics for checking the hypothesis that the number of losses follows a Poisson process is similar to the one used for the severity. Namely, if the counting process of losses follows a Poisson law with intensity λ , then the scaled inter-occurrence times

$$Z_i = \lambda \times (T_i - T_{i-1}), \quad (1.4)$$

Figure 2: Diagnostics of a GPD-Poisson model for BL 2



with $T_0 = 0$, should also be unit exponentially distributed. This could be checked graphically with the corresponding QQ-plot. A scatterplot of the Z_i also gives a notion about the evolution in time. Figure 2 shows some trend in the frequency of occurrences for BL 2. Note that, at the same time, ordered severity residuals don't show any evident trend, hence for BL 2 losses are becoming more frequent without consequent change of the severity. However, joint trends in the severity and frequency can be observed quite often in OpRisk data, and the modeling of joint trends is intensively discussed in the literature, see e.g. (McNeil & Saladin 2000, Schmock 1999).

If some trend in the frequency is detected, then we have to extrapolate it to the future time period over which the aggregate loss has to be considered and to estimate the expected frequency over this period. Hereafter we assume that the constant frequency parameter entering the formulae for the distribution of the aggregate losses is either the estimated constant or the predicted expected value.

1.3 Bayesian inference for mixing Internal and external data

Bayesian estimation has proved to be one of the most efficient methods for parameter adjustment. It is commonly used, when both the common database of observations and a particular series of observations ought to be taken into account. We follow the bayesian methodology for Pareto-type distributions, developed in (Hesselager 1993) and (Reiss & Thomas 1999).

First we repeat the assumptions and notations of the theory based on Bayes' theorem. The basic idea of this approach is to consider the parameter of the distribution as a random variable (if there are several parameters, then we deal with a random vector). Then the *prior distribution* $\pi(\theta)$ is defined as a probability distribution over the space of possible parameter values. Assume that a random variable X has a distribution depending on the parameter vector θ . Then the *model distribution* $f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta)$ is the probability density for the observed sample given a particular value of the parameter (it is identical to the likelihood function). The *posterior distribution* denoted by $\pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x})$ is the conditional probability distribution of the parameters given the observed data.

Using *Bayes' theorem* **the posterior distribution** of the parameters given the observed data is given by

$$\pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) = \frac{f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta)\pi(\theta)}{\int f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta)\pi(\theta)d\theta}. \quad (1.5)$$

Let us consider how this theory can be applied to the problem of estimating severity distributions of operational losses. We should take into account both the internal data from the Bank and the external database, in which the loss data from many banks are accumulated. Values of parameters, estimated from the external data, correspond to the expectation under the prior distribution of the parameter vector while the variance of the prior distribution includes a measure of uncertainty of the estimate. Evaluating the model $f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta)$ at the internal data points corresponding to the selected BL and plugging the result into (1.5), one is able to compute the posterior distribution of the parameters.

Applying this theory to the present problem, we have, first of all, to choose appropriate prior distributions. Under the assumption that the loss severity distribution is GPD and that the loss frequency follows a Poisson process, we need distributions for the shape and scale parameters of the GPD, as well as for the Poisson intensity of the occurrences.

Recall one of the key concepts of the bayesian theory: a prior distribution is said to be a **conjugate prior distribution** for a given model if the resulting posterior distribution is of the same type as the prior. Consider the Pareto distribution with the distribution function given by

$$\tilde{G}_{\xi,\beta}(y) = 1 - \left(1 + \frac{y}{\beta}\right)^{-1/\xi}, \quad \text{for } y > 0 \quad (1.6)$$

(compare this to the GPD distribution (1.1)). Well known results (see, e.g., (Klugman, et al. 2004)) tell us that if the model distribution is the Pareto distribution with the scale

parameter β equal to one, then the gamma distribution for the shape parameter ξ is a conjugate prior. The same is true for the gamma distribution as prior for the Poisson intensity. In this case the estimation of posterior parameters becomes very simple. Unfortunately, for the scaled Pareto distribution with $\beta \neq 1$, as well as for the GPD given in (1.1), such a convenient choice for the prior distribution of ξ does not exist.

In certain cases the model and the prior distribution can be chosen such that the joint distribution of the parameters (as a product of the likelihood function and the prior distribution) can be split into marginal distributions of each parameter. Then the posterior distribution of each parameter can be calculated directly. This is for example the case when the model distribution is the Pareto (1.6). If, moreover, the prior for the shape parameter is the gamma distribution then an explicit representation can be derived (see (Reiss & Thomas 1999) for details). In most other cases, however, no evident way for an explicit expression for the posterior parameters can be found. In these cases numerical methods such as Monte Carlo Markov chain (MCMC) have to be used to calculate the marginal distributions from the joint distribution. We used MCMC for the calculation of the posterior distributions for both GPD and Weibull model distributions.

2 Estimation of the aggregate loss distribution and its characteristics

Before we continue with a short description of three methods of loss aggregation, we would like to mention that different issues arise in the context of quantifying operational risk. As mentioned before one can consider the operation loss of each business line split up into various loss event types or one considers the aggregate operational loss of a business line. Certainly an analysis on the level of event types matters in special in-depth operational risk analyses such as the question of buying insurance against losses in specific business lines due to specific event types. But in the context of quantifying the regulatory capital under the Basel II guidelines it is not required to distinguish losses by their type. We therefore present methods and results on a business line basis. Obviously the methods can be applied on the basis of event types as well.

We will use the following notation in the description of the aggregation methods:

L_i : The aggregate operational loss of business line $i, i = 1, \dots, m$.

N_i : The number of operational losses in BL $i, i = 1, \dots, m$.

$L_{i,n}$: The size of the n_{th} operational loss in BL $i, i = 1, \dots, m$.

2.1 Monte Carlo simulations for aggregating operational risk

Monte Carlo simulation is widely used for operational risk aggregation. Due to its technical simplicity, it plays an important role in the problem of calculating aggregated risk and its

results can be compared to the ones obtained by other methods. For each business line $i = 1, \dots, m$ we proceed as follows:

- Simulate n (say, $n = 10^J$, $J \geq 4$) yearly losses, for short we denote them by $\widetilde{L}_1, \dots, \widetilde{L}_n$, by the following procedure:
 - Simulate the number of loss occurrences in this business line i.e. a realization of N_i .
 - Subsequently simulate the corresponding loss sizes from the distribution of $L_{i,1}$ (we choose $L_{i,1}$ as a prototype of the independently identically distributed losses) for each of these occurrences.

Dealing with heavy-tailed distributions, one usually needs quite a large number of simulations (say, 10^5 simulations for each BL).

- Put the obtained sample in increasing order to get the order statistics $\widetilde{L}_{1:n} \leq \dots \leq \widetilde{L}_{n:n}$, where $\widetilde{L}_{1:n}$ denotes the smallest of the n simulations and $\widetilde{L}_{n:n}$ the biggest simulated loss.
- The element at position αn of the ordered sample is an estimator of the quantile to the level α for the corresponding business line - thus an estimator of VaR to the level α (e.g. choose $\alpha = 0.999$).

The estimation of high quantiles and thus also of the VaR by Monte Carlo techniques suffers from high variances. Monte Carlo techniques nevertheless provide a higher degree of flexibility when complicated structures appear, e.g. in the framework of insurance problems. However, FFT and recursive procedures, as deterministic methods, give in general more reliable results for comparably simple tasks such as calculating the VaR.

2.2 Fast Fourier Transformation

Another way to calculate the distribution of the aggregate loss is to consider its characteristic function (see appendix A for details). While performing a series of convolutions of the severity distribution is clearly not feasible, under the assumption of independence between loss occurrences and loss sizes one can easily calculate the characteristic function (chf) of the aggregate loss. Fix a business line i , then formula (A.5) for the chf of the aggregate loss for BL L_i becomes

$$\widehat{f}_{L_i}(u) = P_{N_i}(\widehat{f}_{L_{i,1}}(u)), \quad (2.1)$$

where $L_{i,1}$ is a random variable modeling the loss size if an operational loss in BL i happens no matter of which type. If sufficiently many data are collected then this procedure can be done for certain event types as well.

Hence, an obvious approach to compute the pdf of the aggregate loss is to calculate the chf (2.1) and then to invert it numerically using the discrete Fourier transformation (DFT). Note that in many cases such as GPD the chf of loss sizes $\widehat{f}_{L_{i,1}}(u)$ can not be calculated explicitly, but it can be computed via DFT as well. An efficient algorithm is provided by the Fast Fourier transformation (FFT) which is available in most statistics or mathematics computer packages.

Thus, the scheme for calculating the density of the aggregate loss is the following:

- Choose an equidistant grid of points at which you want to approximate the loss density say $x_0 \leq \dots \leq x_{n-1} := q_M$ where q_M denotes some upper bound.
- Calculate the density of the severity distribution at these points $f_{L_{i,1}}(x_j)$ for $j = 1, \dots, n$.
- From this sequence, compute the sequence \widehat{f}_l for $l = 0, \dots, n - 1$ using the FFT:

$$\widehat{f}_l = \sum_{k=0}^{n-1} e^{\frac{2\pi i}{n}kl} f(x_k). \quad (2.2)$$

- Plug the result into the expression for the chf of the aggregate loss and perform the inverse Fourier transform in order to calculate an approximation of the density of the aggregate loss on this grid.

Using this approximate density we get an approximate quantile and thereby the value-at-risk.

Remark 2.1 (Truncating severity distributions). *Dealing with heavy-tailed distributions such as GPD, one faces the problem of properly choosing the right endpoint q_M , up to which one considers the loss size distributions. Obviously, fixing n and choosing q_M too high causes a coarser grid leading to an increase of the discretization error. On the other hand q_M should be chosen high enough in order to capture the desired quantile, e.g. in our case the aggregate loss cdf at q_M must exceed 0.999.*

Remark 2.2 (Reducing the aliasing error). *We use aliasing reduction techniques for increased precision of the FFT as described in Section A of the appendix. The method with an exponential window is used in order to calculate a benchmark for the results obtained with the other methods.*

2.3 Applying the Credit Risk⁺ Methodology to Operational Risk

We are passing to the description of another methodology for efficient loss aggregation. The CreditRisk⁺ methodology was developed by Credit Suisse First Boston (Credit Suisse First Boston 1997) for large credit risk portfolios and a detailed description of various extensions can be found in (Gundlach & Lehrbass 2003). This approach was initially developed for

the estimation of aggregate credit risks, and adjusted to operational risk (see for example (Schmock 2006)). Basically it is a compound Poisson model where gamma distributed risk factors scale default intensities (in this context rather: intensities of loss occurrence). One specialty of this approach is, that the calculation of the loss distribution is performed by the expansion of the probability generating function (pgf) of the loss which can be calculated in closed form. In particular the coefficients of the power series expansion of the pgf are calculated as they coincide with the corresponding probabilities. Note that the moment generating function or the characteristic function can be obtained in closed form as well. However, the expansion of the pgf arises naturally as the power series expansion of an analytical function. See appendix A for more information on generating functions.

In order to reduce computational effort losses are counted in integer multiples of a chosen loss unit (this can be one Euro or 100,000 depending on the need for speed and accuracy). Consequently the loss sizes in each business line $(L_{i,n})_{n \in \mathbb{N}}$ for $i = 1, \dots, m$ are assumed to be discrete i.i.d. random variables. This does not limit the choice of the loss size distribution as we can use a discretization of any continuous distribution. Note that there are discretization steps in the FFT approach of Section 2.2 as well. In order to arrive at closed form expressions we furthermore assume the loss sizes to be independent of all other random variables. Finally we will use the probability generating function of the claim size, denoted by $P_{L_{i,1}}(z) = \mathbf{E}[z^{L_{i,1}}]$, which is defined at least for $|z| \leq 1$.

In this description we stick to the original CreditRisk⁺ model with a gamma distributed risk factor. The set-up which we use to calculate the distribution of the operational loss of one particular business line is the following: Consider a risk factor Λ that randomly scales loss intensities. This risk factor is assumed to be gamma distributed such that $\mathbf{E}[\Lambda] = 1$ and $\text{Var}(\Lambda) = \sigma^2 \geq 0$. We assume that, knowing the risk factor, the number of loss occurrences N_i is Poisson distributed with an intensity scaled by the risk factor, i.e.

$$\mathcal{L}(N_i|\Lambda) = \text{Poisson}(\lambda_i \Lambda),$$

where λ_i denotes the expected number of operational losses in business line i for $i = 1, \dots, m$. Integrating over the gamma distribution we arrive at a negative binomial distribution with $\mathbf{E}[N_i] = \lambda_i$ and $\text{Var}(N_i) = \lambda_i(1 + \sigma^2 \lambda_i) = \mathbf{E}[N_i](1 + \sigma^2 \mathbf{E}[N_i])$. Due to the risk factor the number of losses in a business line can be modeled over-dispersed with a degree of over-dispersion depending on σ^2 . The case $\sigma^2 = 0$ simply means that we have a Poisson distribution where the expected number of losses equals its variance. Using the formula for compound Poisson distributions (A.4) the conditional pgf of the loss L_i of business line i is found to be

$$\mathbf{E}[z^{L_i}|\Lambda] = \exp(\Lambda \lambda_i (P_{L_{i,1}}(z) - 1)). \quad (2.3)$$

Using the formula for the moment generating function of the gamma distribution the expectation of (2.3) is given by

$$P_L(z) = (1 - \sigma^2 \lambda_i (P_{L_{i,1}}(z) - 1))^{-\frac{1}{\sigma^2}}. \quad (2.4)$$

Recursive algorithms described in (Haaf, et al. 2003, Schmock 2006) can be applied to calculate the probability distribution of the aggregate loss from this expression.

Random sums $L_i = L_{i,1} + \dots + L_{i,N_i}$ as described above are called compound sums in the insurance literature. Note that the loss distribution in compound Poisson and compound negative binomial models can alternatively be calculated using the Panjer recursion (see, for example, (Klugman et al. 2004)). Although some authors argue that this algorithm is not numerically stable, it is nevertheless shown (Panjer & Wang 1993) that the Panjer recursion in the case of the Poisson as well as in the case of the negative binomial distribution is stable with respect to the relative rounding error. We implemented the Panjer recursion as well to check results obtained by the CreditRisk⁺ algorithm.

The problem of estimating VaR remains sensible approaching this problem with CreditRisk⁺. With the same idea as in the context of FFT, providing severity data for CreditRisk⁺, one must again choose some sufficiently high loss value where the estimated distribution is truncated. In the recursive algorithm for the calculation of the loss distribution only the probabilities of lower losses are needed. Thus, besides unavoidable rounding errors the distribution and thereby the quantiles are calculated exactly up to the truncation point.

3 Description of results, comparison and analysis

In this chapter we present numerical results of the VaR calculation for all business lines. Since confidentiality does not allow us to present real values, absolute values of *all* results are fictitious, but the relative proportions are realistic.

3.1 Overview of results

Due to the *Basel II* regulatory rules (Basel Committee of Banking Supervision 2001), the basic recommendation for the OpRisk capital charge calculation is given by

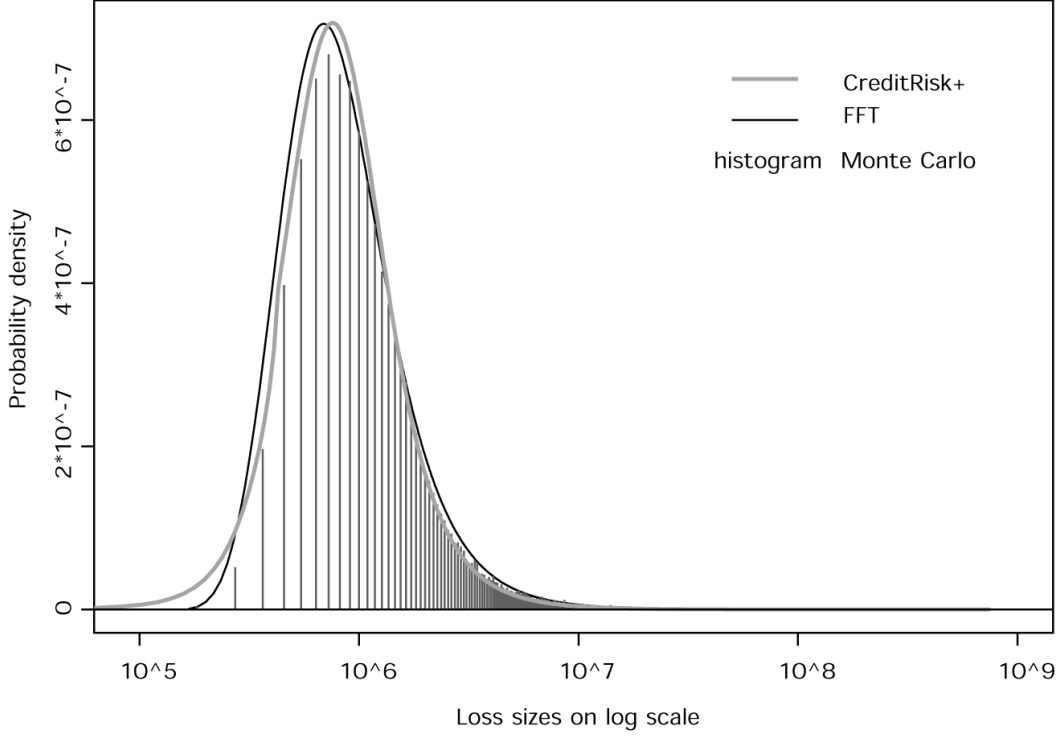
$$C = \sum_{i=1}^m \text{VaR}_\alpha(L_i),$$

where $\text{VaR}_\alpha(L_i), i = 1, \dots, m$ denotes the value-at-risk of the loss of the corresponding BL to the level α . The quantile level used is $\alpha = 99.9\%$, which corresponds to the usual capital charge under the Advanced Measurement Approach of the Basel II guidelines.

Hence, the calculation of regulatory capital for operational risk does not require splitting data by ET's, but only by BL's, which simplifies the problem. Note that the lines of business are encoded as follows: corporate finance (BL 1), trading and sales (BL 2), retail banking (BL 3), commercial banking (BL 4), clearing (BL 5), agency services (BL 6), asset management (BL 7), retail brokerage (BL 8), and private banking (BL 9).

Aggregate loss densities and VaR A graphical illustration of the results using the three methods is given in Figure 3. The distributions obtained using MC, FFT and

Figure 3: Density plots of the aggregate loss of BL 1 obtained by the three methods.



CreditRisk⁺ are relatively close, but obviously the estimated distribution using MC deviates most. A visible difference between FFT and CreditRisk⁺ in the region of small losses can be explained by the fact, that the loss unit used in CreditRisk⁺ is significantly larger than the grid width used in FFT. In particular, for the plot presented the loss unit is 50,000 and the grid width is equal to $q_M/n = 10^9/2^{19} = 1907.4$ (where n is the number of grid points and q_M denotes the highest loss considered).

In Table 1, we present VaR results, obtained with the three methods and pass to the further discussion and comparison of the methods in the following paragraph. Confidence intervals for the value-at-risk are presented in Section B of the appendix.

Business lines 3, 8 and 9 (retail banking, retail brokerage and private banking) are put together due to the advice by experts of the Bank who consider the mechanism of loss generation in these three BL's to be very similar. For BL's 2, 4, 6 and 7 external data were used to obtain the prior estimates of the severity parameters first, and then bayesian inference was applied to adjust the parameters with respect to the internal data according to the methodology described above (see Section 1.3). In order to model losses below the thresholds which were used to fit the severity parameters, empirical distributions based on

	Severity	Frequency*	M-C	FFT	CRP
Line 1 GPD-NB	$\mu = 3500, \xi = 1.12$ $\beta = 7460$	(33.07, 0.9)	662	656.12	656.45
Line 2 W-P	$\mu = 5 \cdot 10^5, \xi = 0.61$ $\beta = 3.13 \cdot 10^6$	44.7	67.4	68.34	68.29
Lines 3,8,9 GPD-NB	$\mu = 7500, \xi = 0.71$ $\beta = 7375$	(23.69, 2.9)	33.5	32.33	32.6
Line 4 W-P	$\mu = 4 \cdot 10^5, \xi = 0.52$ $\beta = 1.38 \cdot 10^6$	7.1	26	27.3	27.26
Line 6 GPD-P	$\mu = 10^4, \xi = 1.15$ $\beta = 18100$	2.2	108.5	110.17	110.31
Line 7 GPD-P	$\mu = 10^4, \xi = 1.2$ $\beta = 15600$	3.2	212	209.47	209.76

Table 1: Fitted parameters and VaR for various business lines. The entries in the first column tell the models used for the severity, as well as for the frequency, where "W" stands for Weibull distribution, "P" for Poisson distribution and "NB" for negative binomial distribution

the internal data were used for all BL's (which is of special importance for BL 2 and BL 4 due to relatively high thresholds).

3.2 Comparative analysis

In this section we illustrate the performance of different methods and compare them on the example of BL 1 (Corporate finance). Thus severity is modeled by a GPD with parameters $\xi = 1.12$, $\mu = 3500$, and $\beta = 7460$. For the occurrences we use a Poisson process with intensity $\lambda = 28.4$.

We compare the results obtained using the three methods with the most precise result, which was obtained using advanced FFT techniques, i.e. FFT with an exponential window. Particularly, with the parameter for the exponential window adjusted to the optimal value, the VaR of the aggregate loss is calculated to be 650.7 million.

Monte Carlo Simulations were done using the *InsightfulTM S-Plus 6.0* package. We denote the sample size by N_{MC} and consider how long it takes to simulate a certain number of samples using an Intel Pentium 2.0 GHz processor. As we notice, the time of random sample generation increases intensively with the sample size. The resulting errors for BL 1 in terms of the deviation in percentage are given in the Table 2.

*For the negative binomial case we use the parametrization (r, b) defined according to Remark 1.1.

Fast Fourier Transformation VaR calculations based on FFT were also done in *S-Plus*. N_{FFT} denotes the number of points in the uniform grid used as input for FFT and IFFT. As a tool for reducing the aliasing effect the input sequence is padded by a set of zeros as large as the number of grid points used. Note that N_{FFT} pointed out in the table denotes the number of points in the grid before padding. The truncation point for the severity is chosen to be 10^9 . The relative errors are presented in Table 3 as the deviation from 650.7 million in percentage.

CreditRisk⁺ Using the CreditRisk⁺ methodology the size of the loss unit is the key parameter of performance and accuracy. The smaller the loss unit, the more recursions are performed. The range of the loss size distribution is chosen analogous to FFT and the results are illustrated in Table 4.

N_{MC}	$5 \cdot 10^6$	$5 \cdot 10^5$	10^5
Machine time (sec.)	207	95	23
Error in percentage	1.1%	4.3%	6.9%

Table 2: Accuracy vs. Speed using MC

N_{FFT}	2^{21}	2^{20}	2^{19}	2^{18}
Machine time (sec.)	35	9.5	4.5	2.5
Error in percentage	-0.012%	-0.06%	-0.16%	-0.38%

Table 3: Accuracy vs. Speed using FFT

$LossUnit$	10,000	25,000	50,000	100,000
Machine time (sec.)	686	110	27	6.7
Error in percentage	+0.043%	+0.088%	+0.18%	+0.38%

Table 4: Accuracy vs. Speed using CreditRisk⁺

We note that FFT and CreditRisk⁺ show almost equal precision deviating from our benchmark by $\pm 0.2\% - \pm 0.4\%$ whereas MC can only compete using extremely large samples. When a higher precision is required, the computational time increases slower in the case of FFT. This fact is clear considering the increase of the computational effort. Using a recursive scheme such as CreditRisk⁺ one needs $O(n^2)$ operations to calculate the probability of losing n times the loss unit, which is significantly higher than the effort of FFT of $O(n \log_2(n))$ (with a grid size which is a power of 2). On the other hand one advantage of CreditRisk⁺ is, that besides rounding errors the only source of error is the discretization error whereas using FFT one has to handle the aliasing error additionally as mentioned before in Remark 2.2.

3.3 Conclusions

In the present paper we applied several methods of loss aggregation in order to analyze the operational risk of a financial institution. We focused on the comparison of an application of the FFT algorithm and the CreditRisk⁺ recursion scheme. In the financial services industry Monte Carlo methods are very popular but without managing the stochastic error of the simulation results the outcomes turn out to be significantly less reliable than the outcomes of deterministic methods. The advantages of FFT and CreditRisk⁺ as deterministic methods are confirmed by our analysis. Concerning accuracy the two deterministic methods show to be equally precise whereas FFT is in general quicker than CreditRisk⁺. Whereas identifying the error sources is more delicate using FFT, the main source of error using CreditRisk⁺ is the discretization of the loss size distribution. The choice between the two methods can, therefore, depend of the particular conditions of the risk analysis, e.g.: the number of risk cells to be analyzed or the period of time between the renewal of the data as this may claim a recalculation of the loss distribution, so that the speed of computations becomes important. The CreditRisk⁺ framework allows to incorporate dependence between business lines. Estimating the parameters of this dependence structure and comparing aggregated losses is a topic for further research.

A The probability generating function and the characteristic function of compound distributions

The probability generating function of the compound sum. We shortly recall the derivation of the probability generating function as well as of the characteristic function of a compound sum of the form

$$S = \sum_{i=1}^N X_i, \tag{A.1}$$

where N is the loss counting variable independent of the i.i.d. sequence $(X_i)_{i=1}^{\infty}$. We denote the cdf of the loss sizes by $F_X(x) := \mathbf{P}[X \leq x]$.

For a random variable N taking only nonnegative integer values consider the probability generating function (pgf) $P_N(z) = \mathbf{E}[z^N] = \sum_{n=0}^{\infty} \mathbf{P}[N = n]z^n$ which is defined and analytic at least for $|z| \leq 1$. Considering the power series expansion of this function $P_N(z) = \sum_{n=0}^{\infty} p_n z^n$ one is able to retrieve the distribution $\mathbf{P}[N = n] = p_n$ for $n \geq 0$ by calculating the coefficients of $P_N(z)$. Denoting the pgf of a compound sum of the form (A.1) by $P_S(z)$, considering integer valued loss sizes and using the independence assumption we

find

$$\begin{aligned}
P_S(z) = \mathbf{E}[z^S] &= \sum_{k=1}^{\infty} \mathbf{P}[N = k] \mathbf{E}[z^{X_1 + \dots + X_k}] \\
&= \sum_{k=1}^{\infty} \mathbf{P}[N = k] \prod_{j=1}^k \mathbf{E}[z^{X_j}] \\
&= \sum_{k=1}^{\infty} \mathbf{P}[N = k] P_X(z)^k.
\end{aligned} \tag{A.2}$$

Finally, one arrives at the well known representation

$$P_S(z) = P_N(P_X(z)), \tag{A.3}$$

where obviously $P_N(z)$ is the pgf of the distribution of loss occurrences and $P_X(z)$ analogous for loss sizes. We get

$$P_S(z) = \exp(\lambda(P_X(z) - 1)) \tag{A.4}$$

for Poisson distributed occurrences and

$$P_S(z) = \frac{1}{(1 + b - bP_X(z))^r}$$

in the case of the negative binomial distribution as defined in (1.3).

The characteristic function of the compound sum. Let

$$\widehat{f}_X(u) = \mathbf{E}[e^{iuX}] = \int_{-\infty}^{\infty} e^{iux} dF_X(x), \quad \text{for } u \in \mathbb{R},$$

denote the characteristic function of the loss size distribution and $\widehat{f}_S(u)$ — the chf of the distribution of the compound sum.

Analogously to the representation (A.2), one immediately obtains

$$\widehat{f}_S(u) = P_N(\widehat{f}_X(u)). \tag{A.5}$$

For the occurrences following a Poisson distribution with intensity λ formula (A.5) becomes

$$\widehat{f}_S(u) = \exp(\lambda(\widehat{f}_X(u) - 1))$$

and for the negative binomial distribution with parameter r and b the corresponding expression is

$$\widehat{f}_S(u) = \frac{1}{(1 + b - b\widehat{f}_X(u))^r}.$$

A remark on errors associated to FFT. An efficient way to compute the chf of a compound distribution as given in (A.5) numerically is to use DFT and, particularly, FFT*. The inverse FFT algorithm is used to get an approximation of the density, however, there is a significant specialty dealing with heavy tailed distributions — the so called *aliasing error*.

The origin of the aliasing error is easier to see on an example of the 2-fold convolution. Consider two functions $f(x)$ and $g(x)$, both having positive support on a uniform grid $x_k = k/N$, $k = 0, \dots, N-1$. They define vectors $\{f_k\}$ and $\{g_k\}$, via $f_k = f(x_k)$, $g_k = g(x_k)$. The approximate values of the convolution $c := f * g$ can then be expressed by

$$c_n = \frac{1}{N} \sum_{k=0}^n f_k g_{n-k} \quad \text{for } n = 0, \dots, N-1. \quad (\text{A.6})$$

Applying the *convolution theorem* in the discrete case we get the following quantity

$$\tilde{c}_n = \frac{1}{N} \sum_{l=0}^{N-1} \hat{f}_l \hat{g}_l e^{2i\pi ln/N} = \frac{1}{N} \sum_{k=0}^{N-1} f_k g_{n-k} \quad \text{for } n = 0, \dots, N-1. \quad (\text{A.7})$$

Note that the range of summation in (A.7) is different from the one in (A.6). Indeed, it implies the appearance of negative vector indices $n-k$ in some of the summands. This can be understood as a consequence of the periodicity of the Fourier transformation: For $n-k < 0$, g_{n-k} has to be identified with g_{n-k+N} , as is most easily seen by applying the identity

$$e^{2i\pi l(n-k)/N} = e^{2i\pi l(n-k+N)/N}$$

to (A.7). To avoid ambiguities we may wish to restrict our notation to indices in the range $0, \dots, N-1$. We may then rewrite (A.7) as

$$\tilde{c}_n = c_n + \sum_{k=n}^{N-1} f_k g_{N+n-k}.$$

The difference between c_n and \tilde{c}_n is usually called aliasing error.

Artificially adding zeros to the initial input sequence – so called *padding by zeros* – is, perhaps, the simplest tool for reducing the aliasing error. In the case of the twofold convolution the aliasing error could be completely eliminated. However, for the n -fold convolution one would have to use $\{\tilde{f}_k\}$, $k := 0, \dots, (nN-1)$ with $\tilde{f}_k = f_k$ for $k \leq N-1$ and $\tilde{f}_k = 0$ for $k > N-1$ to eliminate the aliasing error completely. To reduce the aliasing error significantly in the calculation of the compound distribution one would need at least fourfold padding by zeros, as empirical tests (Schaller & Temnov 2007) show.

Another way to reduce the aliasing error in the FFT algorithm is to apply a transformation which decreases the input values at the right-hand end of the grid and which can

*For an overview of FFT as a special algorithm for DFT see, e.g. (Jerri 1992)

be easily inverted. Probably the simplest of such transformations is the exponential one. Obviously, if $f_\tau(x) := f(x)e^{-x/\tau}$ then

$$(f * g)(x) = e^{x/\tau}(f_\tau * g_\tau)(x).$$

Decreasing the values on one hand reduces the aliasing error but can cause other problems as the precision at which very small numbers can be stored is limited. The problem to choose the optimal value for the parameter τ such that the aliasing error is reduced and that simultaneously the small numbers appearing can still be represented in a satisfactory way is solved in (Schaller & Temnov 2007). In the present paper we use the methodology developed to compute the benchmark for the VaR of the compound distributions for each BL.

B Confidence Intervals for maximum likelihood estimators

Confidence intervals in the context of QQ-Plots We consider a simple but illustrative way of constructing confidence intervals in the context of QQ-plots as proposed in (Schaller 2006) and apply it to the estimated severity distributions. Assume that we have chosen some distribution $F(x) = \mathbf{P}[X \leq x]$ for the severity of some business line. Furthermore we consider the ordered sample of losses from the internal loss database $X_{(1:n)} \leq \dots \leq X_{(n:n)}$. Then it is known that the distribution of $F(X_{(k:n)})$ for $k = 1, \dots, n$ is the same as the distribution of the k -th element of an ordered sample of size n of $(0, 1)$ -uniforms $U_k, k = 1, \dots, n$. This sample of uniforms is usually approximated by simply taking the sequence $U_k := \frac{k}{n+1}$ for $k = 1, \dots, n$.

Comparing the ordered sample $X_{(1:n)} \leq \dots \leq X_{(n:n)}$ and quantiles of the fitted distribution $F^{-1}(U_k), k = 1, \dots, n$ shows us how well $F(x)$ fits the data. If $F(x)$ is the true distribution then the plot of the quantiles of $F^{-1}(U_k), k = 1, \dots, n$ against the ordered sample (i.e. the QQ-plot) should result in a straight line. However, statistical fluctuations can lead to a picture showing deviations from this straight line. The statistical task in this context is to test whether these fluctuations are in admissible bounds or whether the empirical quantiles $X_{(1:n)}, \dots, X_{(n:n)}$ differ significantly from theoretical ones. The following procedure described in (Schaller 2006) allows us to calculate confidence intervals for each $X_{(k:n)}$ for $k = 1, \dots, n$. Considering that $\mathbf{P}[F(X_i) \leq q] = q$ for $i = 1, \dots, n$ and $q \in (0, 1)$, we get the following expression for the probability of the k -th order statistic of our sample to be greater than the quantile to some level q :

$$\mathbf{P}[X_{(k:n)} > F^{-1}(q)] = \mathbf{P}[F(X_{(k:n)}) > q] = \sum_{i=0}^{k-1} \binom{n}{i} q^i (1-q)^{n-i} =: \text{Bin}(n, q, k-1).$$

Thus fixing some confidence level p (e.g. $p = 0.95$) we can solve

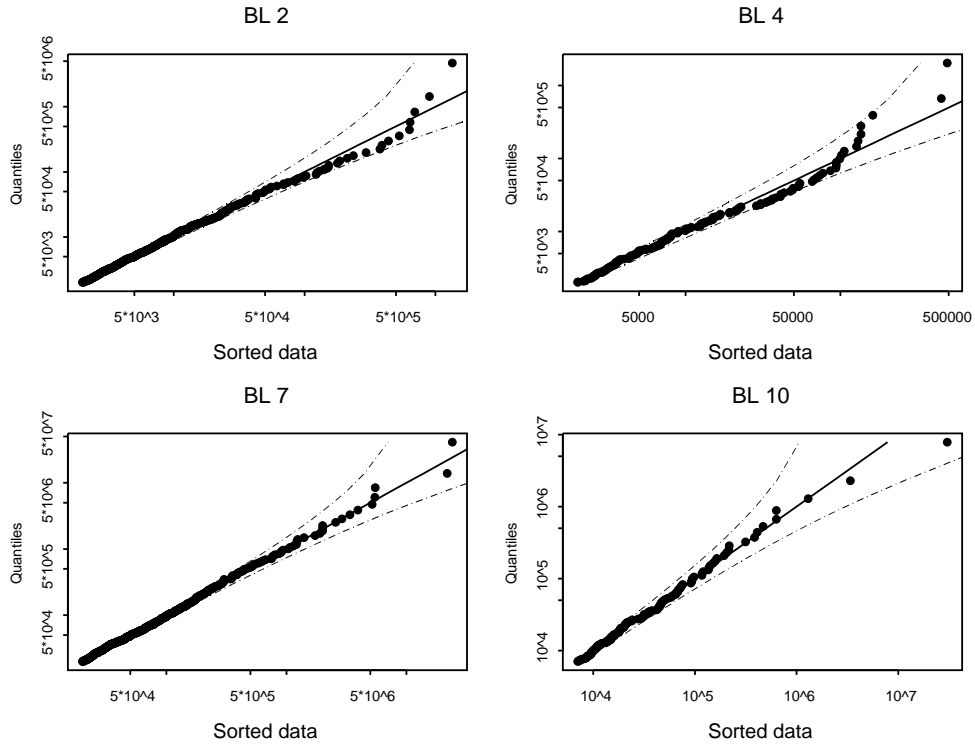
$$\begin{aligned} \text{Bin}(n, q_k^-, k-1) &= p \\ \text{Bin}(n, q_k^+, k-1) &= 1-p \end{aligned}$$

for q_k^- resp. q_k^+ such that the corresponding quantiles fulfil

$$\mathbf{P}[F^{-1}(q_k^-) \leq X_{(k:n)} \leq F^{-1}(q_k^+)] = (1 + p)/2$$

for $k = 1, \dots, n$. Note that q_k^- and q_k^+ only depend on the sample size n and the chosen confidence level p . In Figure 4 we provide a QQ-plot of the severity of losses of selected business lines together with the confidence intervals constructed as above for $p = 0.95$. If points lay outside of these confidence bands we would have to reject the respective distribution. Confidence intervals associated with the largest observations may seem too wide. But considering the heavy-tailedness of the distributions we deal with, it is natural that the largest losses have a wide admissible range in terms of theoretical quantiles.

Figure 4: QQ-plots and confidence intervals.



Confidence intervals according to asymptotic normality We analyze the MLE parameter estimation accuracy for GPD and for Weibull. The problem of calculating approximate confidence intervals for MLE estimators is a classical problem of statistics. Estimating the variance of an MLE in the context of EVT is of particular interest, as the influence of the estimated parameter errors on the VaR is very strong for heavy-tailed distributions.

The precision of MLE estimators for GPD was discussed, e.g., in (Coles 2001, Klugman et al. 2004, Embrechts et al. 1997, Han 2003). The well known way to calculate confidence intervals for MLE estimators using asymptotic normality (see (Klugman et al. 2004) for details) leads to the following expression for a confidence interval to the level 0.95 of the maximum likelihood estimator $\hat{\theta}$ of some parameter θ

$$0.95 = \mathbf{P} \left[-1.96\sqrt{\text{Var}(\hat{\theta})} \leq \theta - \hat{\theta} \leq 1.96\sqrt{\text{Var}(\hat{\theta})} \right], \quad (\text{B.1})$$

where $\text{Var}(\hat{\theta})$ denotes the variance of the estimator which can be estimated by the inverse of the Fisher information matrix (Klugman et al. 2004). Applied to our problem, this method allows us to compute confidence intervals for the estimated parameters for each of the BL's in question. The results are presented in Table 5, as well as estimated lower and upper bounds for VaR corresponding to the uncertainty of the parameters.

Line No	Parameters (shape, scale)	Error bounds	VaR (via FFT)	lower bound	upper bound
Line 1	$\xi = 1.12$ $\beta = 7460$	(0.95, 1.29) (6326, 8594)	656.12	115	3738
Line 2	$\xi = 0.61$ $\beta = 3.13 \cdot 10^6$	(0.65, 0.57) (2.93, 3.13) $\cdot 10^6$	68.34	54.2	88.2
Lines 3,8,9	$\xi = 0.71$ $\beta = 7815$	(0.66, 0.76) (7268, 8362)	32.33	19.2	55
Line 4	$\xi = 0.52$ $\beta = 1.38 \cdot 10^6$	(0.58, 0.46) (1.21, 1.55) $\cdot 10^6$	27.3	18	44
Line 6	$\xi = 1.15$ $\beta = 18100$	(1.07, 1.23) (16833, 19140)	110.17	59.6	203.8
Line 7	$\xi = 1.2$ $\beta = 15600$	(1.1, 1.3) (14352, 16848)	209.47	94	468

Table 5: VaR bounds from confidence intervals

Remark B.1. *One should distinguish the level of the confidence intervals for VaR from the quantile level. The quantile level we use is 99.9% while the level of the confidence intervals is 95%. That is, we consider the VaR for each BL as a random variable due to the parameter uncertainty of the corresponding severity distribution. The choice of the level 95% for confidence intervals is common practice in statistics.*

The lower and upper bounds for VaR displayed in Table 5 are calculated considering the uncertainty of the severity parameters only. In order to determine strict bounds for VaR the uncertainty of the frequency parameters, as well as the error of the procedure of loss

aggregation should also be taken into account. However, as shown above, the error of the loss aggregation is relatively low in comparison to the error stemming from the parameter uncertainty. Concerning the uncertainty of the frequency parameters, one has to keep in mind that a future evolution is extrapolated.

References

- Basel Committee of Banking Supervision (2001). ‘Working paper on the regulatory treatment of operational risk’. available on the web site of the Bank for International Settlements http://www.bis.org/publ/bcbs_wp8.pdf.
- S. Coles (2001). *An introduction to statistical modeling of extreme values*. Springer Series in Statistics. Springer-Verlag London Ltd., London.
- D. R. Cox & E. J. Snell (1968). ‘A general definition of residuals, (with discussion)’. *J. Roy. Statist. Soc. Ser. B* **30**:248–275.
- Credit Suisse First Boston (1997). ‘CreditRisk⁺: A credit risk management framework’. available at <http://www.csfb.com/creditrisk>.
- P. Embrechts, et al. (1997). *Modelling extremal events*, vol. 33 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin.
- R. Fisher & L. Tippett (1928). ‘Limiting forms of the frequency distribution of the largest or smallest number of the sample’. *Proc. Cambridge Philos. Soc.* **24**:180–190.
- M. Gundlach & F. Lehrbass (eds.) (2003). *CreditRisk⁺ in the Banking Industry*. Springer Finance. Springer-Verlag.
- H. Haaf, et al. (2003). ‘Numerically stable computation of CreditRisk⁺’. In (Gundlach & Lehrbass 2003), pp. 69–77.
- Z. Han (2003). *Actuarial modelling of extreme events using transformed generalized extreme value distributions and generalized Pareto distributions*. Ph.D. thesis, The Ohio State University, Ohio, USA.
- O. Hesselager (1993). ‘A class of conjugate priors with applications to excess-of-loss reinsurance’. *ASTIN Bulletin* **23**(1):77–93.
- A. J. Jerri (1992). *Integral and discrete transforms with applications and error analysis*, vol. 162 of *Monographs and Textbooks in Pure and Applied Mathematics*. Marcel Dekker Inc., New York.
- S. A. Klugman, et al. (2004). *Loss models: From data to decisions*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edn.

- A. J. McNeil & T. Saladin (2000). ‘Developing scenarios for future extreme losses using the POT method’. *Extremes and Integrated Risk Management* Published by RISK books, London, available at http://www.math.ethz.ch/~mcneil/pub_list.html.
- H. H. Panjer & S. Wang (1993). ‘On the stability of recursive formulas’. *ASTIN Bulletin* **23**(2):227–258.
- R.-D. Reiss & M. Thomas (1999). ‘A new class of Bayesian estimators in Paretian excess-of-loss reinsurance’. *ASTIN Bulletin* **29**(2):339–349.
- P. Schaller (2006). ‘QQ-plots - error bounds’. workingpaper, Bank Austria - Creditanstalt, Vienna, Austria.
- P. Schaller & G. Temnov (2007). ‘Efficient and precise risk aggregation: applying FFT to heavy tailed distributions’. *Preprint*.
- U. Schmock (1999). ‘Estimating the value of the Wincat coupons of the Winterthur insurance convertible bond: a study of the model risk’. *ASTIN Bulletin* **29**(1):101–163.
- U. Schmock (2006). ‘Modelling dependent credit risks with extensions of CreditRisk⁺ and applications to operational risk (lecture notes)’. Lecture held at the Austrian Central Bank (OeNB) and Vienna University of Technology, available at <http://www.fam.tuwien.ac.at/~schmock/notes/ExtentionsCreditRiskPlus.pdf>.