Estimation of Operational Value-at-Risk with Minimum Collection Thresholds

Anna Chernobai[1], Christian Menn[2], Svetlozar T. Rachev[1,3,*] and Stefan Trück[3]

Working Paper
First Draft: February 18, 2005
This Version: April 20, 2005

Abstract

Due to regulatory capital requirements scheduled to become effective by the end of 2006, financial institutions put substantial effort in collecting loss data and determining adequate loss distributions for operational risk types. However, thresholds in the collection of operational loss data lead to biased estimates as not all loss cases enter internal databases. We provide an approach to the estimation of left-truncated severity data using the EM-algorithm. We extend the model to adjust the frequency and the aggregated loss distributions, and quantify the impact on the Value-at-Risk figures. Furthermore, the paper demonstrates that the effects are more substantial for heavier-tailed distributions. We recommend that financial institutions should consider the effect of collection thresholds both in estimation of the frequency and severity distributions.

[1]Department of Statistics and Applied Probability
University of California, Santa Barbara, CA 93106, USA

[2]School of Operations Research and Industrial Engineering
Cornell University, Ithaca, NY 14853, USA

[3]Institut für Statistik und Mathematische Wirtschaftstheorie
Universität Karlsruhe
Kollegium am Schloss, D-76128 Karlsruhe, Germany

*S. Rachev gratefully acknowledges research support by grants from Division of Mathematical, Life and Physical Sciences, College of Letters and Science, University of California, Santa Barbara, the Deutschen Forschungsgemeinschaft and the Deutscher Akademischer Austausch Dienst.

1 Introduction

A number of unprecedented large losses in the 1990s has directed risk managers' and regulators' attention toward the operational risk. Two frequently cited cases are the bankruptcy of the Barings Bank in 1995 caused by internal fraud, and the 1998 collapse of the Long Term Capital Management (LTCM) due to inadequate business practices. Operational risk has been since acknowledged as a major contributor to banks' and insurance companies' risk positions. Current estimates suggest that the allocation of total financial risk of a bank is done according to 50% credit, market and liquidity 15% and operational risk 35% [5]. For the insurance companies, being affected by their clients' operational risk in addition to their own, adds a considerable task.

The definition of operational risk according to the Basel capital accord is widely accepted [2]:

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

For regulatory purposes, banks are required to follow the Loss Distribution (large banks) or another approach to determine the operational capital charge [1]. The process is structured around the basic elements: internal data, external data, scenario analysis and internal control factors. An accurate data recording environment is crucial for sound modelling. Under a perfect collection process, all events would be detected and duly recorded, and the observed distribution would define the true distribution of the losses. However, the data recording is generally a subject to lower thresholds, with the threshold for data in the European banks' external database being set at around 1,000,000 Euro, while for the internal databases banks normally record data exceeding around 10,000 Euro [2]. As a consequence, loss events smaller in magnitude than 10,000 Euro do not enter the databases. We refer to such data as *missing* data.

This paper analyzes the effects of such missing data on the distributional form of the loss severity. We make use of a modified version of Dempster's *Expectation-Maximization (EM) algorithm* to 'recover' the true Maximum Likelihood parameters of assumed distribution. The EM algorithm was suggested for truncated data in operational risk in [3]. We emphasize that the severity and frequency distributions of the operational risk are inter-related in the sense that if the missing data significantly alters the severity, the frequency distributions also require adjustment. Hence, if the missing data in the lower quantiles of the severity distribution is accounted for, it can prevent over-estimating higher quantiles and hence can avoid an over-stated *Value-at-Risk* (VaR for short) based capital charge. On the other hand, the intensity factor of the frequency distribution would have to be increased by a factor proportional to the 'information loss', i.e. the difference in the expected fraction of data below the threshold, and this has an opposite effect on the capital charge. This paper examines both effects, and determines the aggregated impact on the operational capital charge. We also examine the sensitivity of the estimates with respect to the threshold levels and distributional patterns.

2 Expectation-Maximization Algorithm for Left-Truncated Data

2.1 Collection Thresholds

A complete collection process for operational losses would yield *true distributions* for severity and frequency. However, thresholds in data collection lead to left-truncated *observed distributions* for severities, and a reduced frequency count. This type of missing data is called *nonrandomly missing data*. Threshold adjustments may be of particular interest for operational risk quantification, due to its possible impact on the capital charge: ignoring the missing data would lead to, possibly significant, scaling up of predictions, and an underestimation of frequency would scale them down. Hence, the effect on VaR figures of, on one hand, overestimated severities and, on the other hand, underestimated frequencies, is not unique and ought to be fully examined. Once the truncated data distribution is estimated, there are two distinct approaches for subsequent steps:

1. Work with the *observed* severity distribution, and the *observed* frequencies.

2. Carry out necessary adjustments, and work with the true (or *complete*) severity distribution, and adequately adjust for the frequency count.

Depending on the chosen method, both the true and the observed distributions for frequency and severity could be of interest. The complete data density specification would require the form:

$$f(x \mid \boldsymbol{\theta}^*) = \begin{cases} \frac{f(x|\boldsymbol{\theta}^*)}{\int_0^\infty f(x|\boldsymbol{\theta}^*)dx} & \text{for} \quad x > 0 \\ 0 & \text{for} \quad x \leq 0 \end{cases} \tag{2.1}$$

since the losses are assumed to take strictly positive values. $\boldsymbol{\theta}^*$ denotes the true parameter space for complete data.

In the next section we introduce a modified version of the EM algorithm that we use to 'recover' the complete-data severity distribution.

2.2 General Methodology

The original EM algorithm was fully developed by Dempster *et. al* in 1977 [6]. This iterative procedure consists of two parts - the Expectation step and the Maximization step. In the end, we obtain the *posterior* form of the parameter vector, given the observed data sample and information on the existence of missing data. For our purposes, let the observed (incomplete) data $\tilde{y}$ be a realization from the sample space $\tilde{\mathcal{Y}}$, and the complete-data set $\tilde{x}$ on $\tilde{\mathcal{X}}$ is observed indirectly through $\tilde{y}$. Hence, the mapping $\tilde{\mathcal{Y}} = \mathcal{X}(\tilde{y})$ ensures the map $\mathbf{x} \mapsto \tilde{\mathrm{y}}(x)$ takes place. The sample space $\mathcal{Y}$ of nonrandomly missing data $\mathbf{y}$ is such that

$\mathcal{X} = \mathcal{Y} \cup \tilde{\mathcal{Y}}$. We denote θ^* to be the complete-data parameter space, and θ to be the parameter space under the observed data specification.

It is notable that unlike most problems to which EM algorithm is applied, *we do not know the exact number (or fraction) of the data that is missing.* We only know that the data is missing not at random, but below a pre-specified threshold H (until zero). Hence, we present a modified version of the EM algorithm, and derive the Maximum Likelihood estimates, without the knowledge of the number of missing data points. Because of the number of missing data is unknown, it is not possible to compute the log-Likelihood function for the *entire* complete-data set. However, the expected value of the observed data can be computed iteratively, for each round of the algorithm.

There could be many possibilities for complete-data specification $f(x|\theta^*)$, given the incomplete-data specification $g(x|\theta^*)$:

$$g(y|\theta^*) = \int_{\mathcal{X}(\tilde{\mathbf{y}})} f(x|\theta^*)dx \tag{2.2}$$

Suppose that $\theta^{(k)}$ denotes the value of the parameters obtained on the kth iteration of the algorithm. The Expectation and Maximization steps (E-step and M-step), at each cycle k, are constructed in the following fashion:

E-*step*: Calculate the expected value of the log-Likelihood function of complete-data set, based on the $i = 1, 2, ..., m$ incomplete data points and the parameter values obtained in the kth cycle:

$$t^{(k)} = \mathbb{E}_{\boldsymbol{\theta}^{(k)}}\Big[l(\mathbf{x}|\boldsymbol{\theta})|\tilde{\mathbf{y}}\Big] = \mathbb{E}_{\boldsymbol{\theta}^{(k)}}\Big[l(\mathbf{y}|\boldsymbol{\theta})|\tilde{\mathbf{y}}\Big] + (1 - Q^{(k)})l(\tilde{\mathbf{y}}|\boldsymbol{\theta}) \tag{2.3}$$

where $Q^{(k)}$ denotes the probability that a data point lies below the threshold H and above 0, under the current cycle's parameter set values.

Remark: The density associated with the log-Likelihood function, is restricted (or adjusted) to the non-negative support, as in Equation 2.1.

M-*step*: Maximize the expected log-likelihood from the E-*step* with respect to the parameters, and set them as the new cycle's initial parameter values $\theta^{(k+1)}$:

$$\mathbb{E}_{\boldsymbol{\theta}^{(k)}}\left[l(\mathbf{x}|\boldsymbol{\theta})|\tilde{\mathbf{y}}\right] = t^{(k)}. \tag{2.4}$$

Every cycle of the algorithm produces new estimates of the unknown parameter set, based on the estimates from the previous cycle. Moreover, each iteration cycle increases the likelihood of the sample. The limit of this sequence of estimates leads to a (local) maximum of the log-likelihood function.

2.3 Example with Lognormal Distribution

In this section, for the practical purposes, we illustrate the estimation procedure for the case of the Lognormal distribution. [1]

For distributions that belong to the exponential family, in the E-step it suffices to compute the expected value of the sufficient statistics. Then, these are used to compute the value of the expected log-Likelihood function.

Let the complete-data vector $\mathbf{x} = (x_1, ..., x_N)'$ be a random sample drawn from the Lognormal distribution with parameter set $\boldsymbol{\theta} = \{\mu, \sigma\}$, with N being the total number of observations in the complete data set, m being the number of observed data, and $N-m$ being the number of missing data (which is unknown). The sufficient statistics vector is $T(\mathbf{X}) = \{\sum_{i=1}^{N} log(X_i);\ \sum_{i=1}^{N} log^2(X_i)\}$, and taking expectations for *each* data point, we obtain at each iteration:

$$\mathbb{E}_{\boldsymbol{\theta}^{(k)}}\left[\frac{1}{N}\sum_{i=1}^{N} log(X_i)|\tilde{\mathbf{y}}\right] = (1-Q^{(k)})\frac{1}{m}\sum_{i=1}^{m} log(x_i) + \int_0^H log(x) f(x|\tilde{\mathbf{y}}, \boldsymbol{\theta}^{(k)})dx$$

$$\mathbb{E}_{\boldsymbol{\theta}^{(k)}}\left[\frac{1}{N}\sum_{i=1}^{N} log^2(X_i)|\tilde{\mathbf{y}}\right] = (1-Q^{(k)})\frac{1}{m}\sum_{i=1}^{m} log^2(x_i) + \int_0^H log^2(x) f(x|\tilde{\mathbf{y}}, \boldsymbol{\theta}^{(k)})dx$$

The M-step simplifies to finding the unique solution to the following equation:

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}^{(k)}}\left[l(\mathbf{x}|\boldsymbol{\theta})|\tilde{\mathbf{y}}\right] = \\
&\quad -(1-Q^{(k)})\left[\frac{m\log(2\pi\hat{\sigma}^2)}{2} + \sum_{i=1}^{m} log(x_i) + \frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{m}[log(x_i) - \hat{\mu}^{(k)}]^2\right] - ... \\
&\quad -Q^{(k)}\frac{(N-m)\log(2\pi\hat{\sigma}^2)}{2} - (N-m)\left[\int_0^H \log(x) f(x|\tilde{\mathbf{y}}, \boldsymbol{\theta}^{(k)})dx + ... \right.\\
&\quad \left. +\frac{1}{2\hat{\sigma}^2}\int_0^H [log(x) - \hat{\mu}]^2 f(x|\tilde{\mathbf{y}}, \boldsymbol{\theta}^{(k)})dx\right]
\end{aligned}$$

which, alternatively, implies the likelihood for *each* data point of the form

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}^{(k)}}\left[l(x_i|\boldsymbol{\theta})|\tilde{\mathbf{y}}\right] = \\
&\quad -(1-Q^{(k)})\left[\frac{\log(2\pi\hat{\sigma}^2)}{2} + \frac{1}{m}\sum_{i=1}^{m} log(x_i) + \frac{1}{2m\hat{\sigma}^2}\sum_{i=1}^{m}[log(x_i) - \hat{\mu}]^2\right] - ... \\
&\quad -Q^{(k)}\frac{\log(2\pi\hat{\sigma}^2)}{2} - \int_0^H \log(x) f(x|\tilde{\mathbf{y}}, \boldsymbol{\theta}^{(k)})dx - ... \\
&\quad -\frac{1}{2\hat{\sigma}^2}\int_0^H [log(x) - \hat{\mu}]^2 f(x|\tilde{\mathbf{y}}, \boldsymbol{\theta}^{(k)})dx, \quad i = 1, 2, ..., N
\end{aligned}$$

[1] BIS suggests using Lognormal distribution to model the severity distribution of operational losses [1].

The$(k+1)$th step's posterior Maximum Likelihood estimation of μ and σ^2 yields the new estimates:

$$\widehat{\mu}_{\text{MLE}}^{(k+1)} = (1 - Q^{(k)})\frac{1}{m}\sum_{i=1}^{m}\log(x_i) + \int_0^H \log(x)f(x;\boldsymbol{\theta}^{(k)})dx$$

$$\widehat{\sigma^2}_{\text{MLE}}^{(k+1)} = (1 - Q^{(k)})\frac{1}{m}\sum_{i=1}^{m}\log^2(x_i) + \int_0^H \log^2(x)f(x;\boldsymbol{\theta}^{(k)})dx - \widehat{\mu^2}_{\text{MLE}}^{(k+1)}$$

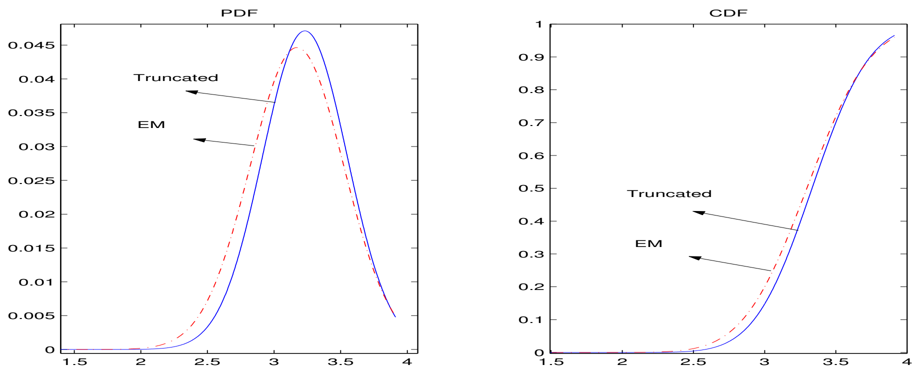


Figure 1: Comparison of PDF and CDF for the example. The bold graph is the original fit. The dotted graph is based on the EM algorithm.

Table 11 illustrates the convergence results with the EM algorithm for a simple problem with five observed values $\tilde{y} = \{20, 23, 25, 30, 50\}$ and a threshold level $H = 15$. As the start values for the algorithm we chose the MLE estimates of the unconditional Lognormal density, $\mu_0 = 3.3327$ and $\sigma_0^2 = 0.1011$. This yields the value of the log-Likelihood for the unconditional Lognormal distribution applied to the observed data of -18.0299. The algorithm iterates until $\delta = \max_k \{\mu^{(k)} - \mu^{(k-1)}, \sigma^{2(k)} - \sigma^{2(k-1)}\} < 0.1 \cdot 10^{-15}$, i.e. until the MLE estimates converge (the value is chosen arbitrarily). The posterior MLE estimates appear in the bottom row of Table 11. The last column shows the log-Likelihood values for the observed portion of the data.

This simple example illustrates how the parameters for the assumed Lognormal distribution were not appropriate for the complete sample, once the existence of some missing data is recognized. It is viewed from the increase in the scale parameter, and a drop of the location parameter. The log-Likelihood function of the observed data when missing data is accounted for, is also increased. Figure 1 illustrates the difference in fitted densities and distributions. In the next section we will discuss robustness of the Maximum Likelihood estimators under the EM algorithm.

k	$\mu^{(k)}$	$\sigma^{2(k)}$	$logL_{obs}^{(k)}$
1	3.332665	0.1011401	-17.58342
2	3.314242	0.1129907	-17.40839
3	3.305652	0.1182587	-17.33105
4	3.301268	0.1209048	-17.29199
5	3.29894	0.1223007	-17.27129
6	3.29768	0.1230544	-17.26009
7	3.29699	0.1234662	-17.25396
8	3.29661	0.1236928	-17.25059
9	3.2964	0.1238178	-17.24873
10	3.296284	0.1238869	-17.24770
11	3.29622	0.1239251	-17.24713
12	3.296185	0.1239463	-17.24681
13	3.296165	0.123958	-17.24664
14	3.296154	0.1239645	-17.24654
15	3.296148	0.1239681	-17.24649
16	3.296145	0.1239701	-17.24646
17	3.296143	0.1239712	-17.24644
18	3.296142	0.1239719	-17.24643
19	3.296141	0.1239722	-17.24643
20	3.296141	0.1239724	-17.24642
21	3.296141	0.1239725	-17.24642
22	3.296141	0.1239725	-17.24642
23	3.29614	0.1239726	-17.24642
$\vdots$	$\vdots$	$\vdots$	$\vdots$
55	3.29614	0.1239726	-17.24642

Table 1: Iteration results of EM-algorithm for the example. Convergence was achieved after 55 iterations.

2.4 Robustness of the EM Estimators

2.4.1 Properties of the EM Estimators

Here, we briefly present major properties of the EM estimators. Suppose that $\boldsymbol{\theta}^{(k)}$ converges to $\boldsymbol{\theta}^*$ during the kth iteration. Then the following properties hold [6].

Property 1. *For any $\boldsymbol{\theta} \in \Theta$, expected log-Likelihood of the missing data is maximized at each iteration under the new parameter estimates:*

$$\mathbb{E}_{\boldsymbol{\theta}^{(k)}}\left[l(\mathbf{y}|\boldsymbol{\theta})|\tilde{\mathbf{y}}\right] \leq \mathbb{E}_{\boldsymbol{\theta}^{(k)}}\left[l(\mathbf{y}|\boldsymbol{\theta}^{(k)})|\tilde{\mathbf{y}}\right] \tag{2.5}$$

Property 2. *The EM algorithm increases the log-Likelihood of observed-data sample at each iteration:*

$$l(\tilde{\mathbf{y}}|\boldsymbol{\theta}^{(k+1)}) \geq l(\tilde{\mathbf{y}}|\boldsymbol{\theta}^{(k)}) \tag{2.6}$$

with equality iff

$$\mathbb{E}_{\boldsymbol{\theta}^{(k)}}\left[l(\mathbf{x}|\boldsymbol{\theta}^{(k+1)})|\bar{\mathbf{y}}\right] = \mathbb{E}_{\boldsymbol{\theta}^{(k)}}\left[l(\mathbf{x}|\boldsymbol{\theta}^{(k)})|\bar{\mathbf{y}}\right] \tag{2.7}$$

The first property guarantees the convergence. The second property ensures that the log-Likelihood of the observed data is increased with the progress of the algorithm.

2.4.2 Numerical Comparisons

In this section we show how the EM algorithm can be used in order to effectively unveil the true parameters for the complete-data, given nonrandomly missing data and given the parameters from the distribution fitted to the observed-data sample. An important assumption is made that both observed-data sample and the missing data portion follow the same distribution.

We present the results of several simulations. We first generated 100 samples of 1000 data points from various distributions. The parameters of assumed distributions were first computed with the MLE estimators, using the complete-data sample. These are denoted with the subscript in the footage of the following tables. Then, the samples were truncated at various thresholds, and the data below the threshold would be taken as missing. This section presents exemplary results for the Exponential with $\lambda = 0.001$ and Lognormal distribution with $\mu_0 = 5$, $\sigma_0^2 = 2$. The first columns in Table 2 and 3 show the cut-off level. The thresholds were chosen at approximately 10%, 20%, 40% and 60%. The table further demonstrate the parameter estimates for an unconditional estimation procedure and the estimates obtained using the EM algorithm. We introduce the following notations:

> Q^0: true proportion of the data below the threshold, based on the true distribution,
>
> Q^{Tr}: estimated proportion of the data below the threshold, when unconditional density is fitted to the truncated data (Trunc.),
>
> Q^{EM}: estimated proportion of the data below the threshold, when the EM algorithm is performed on the truncated data.

Similarly, the superscripts (subscripts) 0, Tr, EM attached to the parameters, denote the state of the data to which assumed distribution was fitted. The Mean Square Error (MSE) based on 100 samples is calculated as $MSE(\hat{\theta}^{EM}) = \frac{1}{100}\sum_{i=1}^{100}(\hat{\theta}_i^{EM} - \theta_0)^2$.

Table 3 presents the mean values for the parameters and percentages (Q) based on 100 samples of 1000 data points in each. The numbers in the percentages refer to the fraction of the estimated parameter value (or percentage below the threshold) to the real figure. The table demonstrates that EM algorithm allows to almost exactly retrieve the true parameters of the complete-data sample.

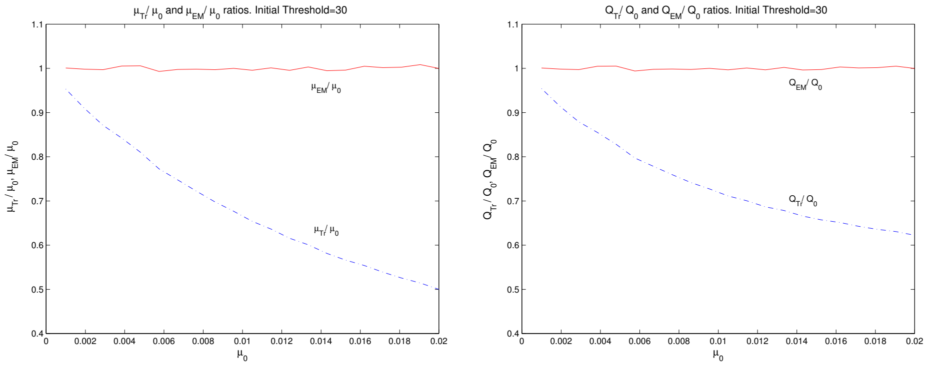
Exponential distribution: $f_X(x;\lambda) = \lambda e^{-\lambda x}$, $\lambda_0 = 0.001$							
Truncation (Q^0)	m	$\boldsymbol{\theta}$	*Trunc.* (θ^{Tr}/θ_0)	*EM* (θ^{EM}/θ_0)	*MSE*	Q^{Tr} (Q^{Tr}/Q^0)	Q^{EM} (Q^{EM}/Q^0)
$H = 110$ (10.4%)	1000	λ	0.000897 (0.8968)	0.000995 (0.995)	$7.45 \cdot 10^{-10}$	9.39% (0.902)	10.37% (0.995)
$H = 250$ (22.1%)	1000	λ	0.000800 (0.8004)	0.0010 (1.0008)	$6.39 \cdot 10^{-10}$	18.14% (0.820)	22.13% (1.001)
$H = 500$ (39.4%)	1000	λ	0.000668 (0.6684)	0.0010 (1.0040)	$4.90 \cdot 10^{-10}$	28.41% (0.722)	39.46% (1.003)
$H = 1000$ (63.2%)	1000	λ	0.000499 (0.4993)	0.0010 (0.9978)	$9.91 \cdot 10^{-10}$	39.30% (0.62)	63.11% (0.998)

Table 2: EM-algorithm with data generated from Exponential distribution, $\lambda_0 = 0.001$. Figures are based on 100 samples of 1000 points after truncation.

Lognormal distribution: $f_X(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}x} exp(-\frac{(\log(x)-\mu)^2}{2\sigma^2})$, $\mu_0 = 5,\ \sigma_0^2 = 2$							
Truncation (Q^0)	m	$\boldsymbol{\theta}$	*Trunc.* (θ^{Tr}/θ_0)	*EM* (θ^{EM}/θ_0)	*MSE*	Q^{Tr} (Q^{Tr}/Q^0)	Q^{EM} (Q^{EM}/Q^0)
$H = 30$ (12.9%)	1000	μ	5.3272 (1.065)	4.9669 (0.993)	0.0039	4.79% (0.3709)	13.59% (1.0527)
		σ^2	1.3373 (0.669)	2.0314 (1.016)	0.0209		
$H = 50$ (22.1%)	1000	μ	5.5382 (1.108)	4.9889 (0.9978)	0.0077	6.28% (0.2845)	22.42% (1.0150)
		σ^2	1.1275 (0.564)	2.0202 (1.0101)	0.0218		
$H = 100$ (39%)	1000	μ	5.8819 (1.176)	5.0323 (1.007)	0.0279	8.13% (0.2084)	37.64% (0.9650)
		σ^2	0.8369 (0.419)	1.9230 (0.962)	0.0811		
$H = 200$ (58.4%)	1000	μ	6.3150 (1.263)	4.9959 (0.999)	0.0622	9.96% (0.1707)	58.01% (0.9942)
		σ^2	0.6276 (0.3138)	1.9681 (0.9840)	0.0786		

Table 3: EM-algorithm with data generated from Lognormal distribution, $\mu_0 = 5,\ \sigma_0^2 = 2$. Figures are based on 100 samples of 1000 points after truncation.

This is based on important assumption that the assumed distribution is the true distribution for all data points. As expected, the Mean Square Errors increase as the threshold level increases, which accounts for a higher degree of uncertainty. The difference between the last two columns concludes what fraction of the data below the threshold has been ignored under the truncation. It increases proportionally with the threshold level. It is notable also, that for the lognormal distribution with a heavier mass in the tails (see the Appendix), these gaps are more significant than for the lighter-tailed ones: for heavier tailed distributions the true fraction of data below the threshold is highly under-estimated. the effect becomes stronger the higher the threshold is chosen.



(a) λ_{Tr}/λ_0, λ_{EM}/λ_0 (b) Q_{Tr}/Q_0, Q_{EM}/Q_0

Figure 2: Robustness check for Exponential distribution. Threshold = 30. A sample of 100,000 points is above each cutoff level.

	λ_0				
	0.001	0.005	0.010	0.015	0.020
Q_0	2.96	13.93	25.92	36.24	45.12

Table 4: Quantiles (%) for Exponential example of Figure 2.

$\sigma_0^2 \setminus \mu_0$	4	5	6	7	8
1.5	47.14	18.72	4.41	0.58	0.00042
2.0	47.52	22.09	6.99	1.45	0.19
2.5	47.78	24.57	9.33	2.54	0.49
3.0	47.97	26.50	11.40	3.73	0.91
3.5	48.12	28.04	13.22	4.94	1.44
4.0	48.25	29.32	14.82	6.13	2.05
4.5	48.35	30.40	16.25	7.27	2.70
5.0	48.43	31.33	17.52	8.36	3.38

Table 5: Quantiles (%) for Lognormal example of Figure 3.

Figure 3 further explore the robustness properties for the EM estimators. The figures are based on the Lognormal example, but the principles hold for other distributions. The top row of figure 3 demonstrates how the ratio of the EM-based scale parameter σ^2_{EM} to the true scale parameter σ^2_0, is affected by varying the threshold level (in the absolute terms) and the initial scale parameter, keeping the location parameter fixed. In the ideal case, it should equal unity. Since the cutoff level in the percentage terms is affected when parameters are changed, we present the table (right) which demonstrates the effect. In the middle row of Figure 3 a similar procedure is demonstrated for the location parameter ratio, with fixed scale parameter. It is clear from the pictures and corresponding tables, that *for the threshold levels roughly under 50%*, the EM-based estimates of both μ and σ^2 stay within 20% of the initial (true) parameter values. Notably, even for very high thresholds (above 50%) the parameter values obtained from the EM procedure remain still very close to the true values. The bottom row of Figure 3 portrays the same issue from yet another angle: we fix the absolute threshold level, and change the initial parameter values, and look at how the EM-estimated fraction of observations under the threshold is close to the true fraction. Again, as long as the cutoff level is roughly under 50%, the ratio of the EM-based fraction to the true fraction is very close to unity.

In the following, empirical, section of the paper, we will use the results and apply the procedure to the real operational risk data. The EM algorithm would be used to reveal the true parameters of the complete-data for several business lines in our data set.

3 Application of EM Algorithm to Operational Loss Data

3.1 Loss Severity Distributions

In this section we apply the EM algorithm to European operational risk data. The data set consists of the loss severity and frequency data for five types: *"Relationship", "Human", "Processes", "Technology"* and *"External"*. The results of fitting various unconditional distributions to the losses were presented in [4]. The minimum values of most of the five data sets were slightly under 10,000 USD, however only a few data points were of low magnitude for the five data sets - since the data sets come from an external database, low magnitude losses were in general not recorded. To ensure comparability, we truncated all data sets strictly at 1000,000 USD, which is a standard set by the Basel Committee for the use of external data bases.

Thus, in our analysis we assume the data to stem from an external loss database. The parameters of various distributions are not representative of the behavior of the losses of each financial institution. We would like to emphasize, that this section purely demonstrates the technique. However, in practice, if the data is truncated at a certain threshold we encourage banks to apply the same technique to their *internal* databases.

Fitted Exponential distribution							
Loss type	$\boldsymbol{\theta}$	*Trunc.* $\times 10^{-10}$	*EM* $\times 10^{-10}$	$logL^{Tr}$	$logL^{EM}$	Q^{Tr}	Q^{EM}
Relationship	λ	98.56	99.54	-10067.44	-9967.75	0.98%	0.99%
Human	λ	56.77	57.09	-11032.78	-10969.98	0.566%	0.569%
Processes	λ	27.05	27.13	-4332.17	-4320.44	0.270%	0.271%
Technology	λ	110.80	112.00	-1004.56	-993.37	1.10%	1.11%
External	λ	93.37	94.25	-3722.45	-3687.54	0.93%	0.94%

Table 6: Distributional adjustment for operational losses with fitted Exponential distribution, with 1,000,000 USD truncation.

Fitted Lognormal distribution							
Loss type	$\boldsymbol{\theta}$	*Trunc.*	*EM*	$logL^{Tr}$	$logL^{EM}$	Q^{Tr}	Q^{EM}
Relationship	μ	16.7514	16.2544	-9698.77	-8517.55	4.57%	12.47%
	σ^2	3.0234	4.4825				
Human	μ	16.6037	14.9670	-10321.70	-6852.51	7.81%	34.58%
	σ^2	3.8661	8.4294				
Processes	μ	17.8452	17.5631	-4181.66	-3951.12	2.77%	5.60%
	σ^2	4.4264	5.5632				
Technology	μ	16.8985	15.9929	-987.91	-795.40	5.93%	20.00%
	σ^2	3.9031	6.6951				
External	μ	16.5353	15.4631	-3543.27	-2666.87	6.72%	25.44%
	σ^2	3.2995	6.2158				

Table 7: Distributional adjustment for operational losses with fitted Lognormal distribution, with 1,000,000 USD truncation.

We get the following results: fitting a thin-tailed distribution to the data a much lower proportion of data is estimated to be missing. For the exponential depending on the loss type the estimated frequency varies between 0.270% and 1.100% for the unconditional fit and between 0.271% and 1.110% for the conditional fit using the EM algorithm. Further the differences between the estimated parameters and thus, also the estimated fraction of missing data is very small. We conclude that for light-tailed distributions and a comparably high threshold is hardly makes any difference whether the naive or a conditional estimation method is chosen.

The effect becomes much stronger if we assume a heavy-tailed distribution for the data. For example for the lognormal distribution parameters show large differences depending on the chosen estimation technique. We observe a decrease in the location parameter, and an increase in the scale parameter, as one can expect. Further the estimated fraction of missing data compared to the ex-

Fitted Gamma distribution							
Loss type	$\boldsymbol{\theta}$	*Trunc.*	*CMLE*	$logL^{Tr}$	$logL^{CMLE}$	Q^{Tr}	Q^{CMLE}
Relationship	α	0.3921	0.1610	-9835.76	-5832.98	12.74%	41.53%
	β	$2588{\cdot}10^5$	$3689{\cdot}10^5$				
Human	α	0.2908	0.1319	-10529.5	-5911.81	17.26%	44.53%
	β	$6058{\cdot}10^5$	$7414{\cdot}10^5$				
Processes	α	0.3561	0.1665	-4212.63	-2899.25	9.46%	31.85%
	β	$10380{\cdot}10^5$	$15136{\cdot}10^5$				
Technology	α	0.4542	0.1843	-989.24	-624.54	10.19%	37.74%
	β	$1988{\cdot}10^5$	$3052{\cdot}10^5$				
External	α	0.3450	0.1461	-3604.39	-2029.63	15.48%	44.47%
	β	$3105{\cdot}10^5$	$4075{\cdot}10^5$				

Table 8: Distributional adjustment for operational losses with fitted Gamma distribution, with 1,000,000 USD truncation.

Fitted Weibull distribution							
Loss type	$\boldsymbol{\theta}$	*Trunc.*	*CMLE*	$logL^{Tr}$	$logL^{CMLE}$	Q^{Tr}	Q^{CMLE}
Relationship	α	0.00006	0.0032	-9762.03	-9660.40	11.50%	33.72%
	β	0.5502	0.3522				
Human	α	0.00024	0.0221	-10412.94	-10239.29	15.31%	62.80%
	β	0.4723	0.2154				
Processes	α	0.00008	0.0010	-4192.61	-4169.00	7.56%	17.17%
	β	0.5010	0.3805				
Technology	α	0.00003	0.0020	-988.38	-980.37	9.11%	28.21%
	β	0.5792	0.3705				
External	α	0.00013	0.0164	-3570.50	-3519.52	14.32%	51.03%
	β	0.5102	0.2732				

Table 9: Distributional adjustment for operational losses with fitted Weibull distribution, with 1,000,000 USD truncation.

ponential distribution is much higher, depending on the loss type the estimated frequency varies between 2.77% and 7.81% for the unconditional fit and between 5.60% and 34.58% for the conditional fit using the EM algorithm. We also observe a big difference between the estimation fraction of missing data dependent on the chosen estimation technique. The 'information loss' is estimated to be 2 to 5 higher if the conditional estimation method using EM is chosen, as is viewed from the last two columns, which is a significant change. The log-Likelihood function of complete-data set is clearly higher under the EM procedure than if the missing data is ignored and unconditional Lognormal distribution is fitted to observed data.

Similar effects can be observed for the fitted Weibull and Gamma distribu-

Fitted 1-parameter Pareto distribution							
Loss type	$\boldsymbol{\theta}$	*Trunc.*	*CMLE*	$logL^{Tr}$	$logL^{CMLE}$	Q^{Tr}	Q^{CMLE}
Relationship	β	0.0597	0.3406	-10655.18	-9753.10	56.17%	99.10%
Human	β	0.0602	0.3587	-11268.13	-10283.23	56.49%	99.30%
Processes	β	0.0560	0.2482	-4540.94	-4229.94	53.89%	96.76%
Technology	β	0.0592	0.3244	-1077.74	-989.27	55.85%	98.87%
External	β	0.0605	0.3677	-3885.10	-3540.35	56.64%	99.38%

Table 10: EM-algorithm results for operational losses with fitted 1-parameter Pareto distribution, with 1,000,000 USD truncation.

tion. For both distributions the fraction of missing data is clearly estimated to be higher using the conditional approach that also gives higher loglikelihood values for all loss type categories. We should assume that the unconditional approach again clearly underestimates the true number of events.

For the fitted one-parameter Pareto distribution the results are rather extreme, the distribution seems to be too heavy-tailed to make the results reliable. While for the unconditional case a fraction of approximately 55% is estiamted to be missing, the EM algorithm gives parameter estimates indicating a fration of more than 95% to be missing. We consider these results as rather unrealistic and will exclude the most heavy-tailed distribution, the one-parameter Pareto from further analysis, since it seems not adequate for the considered data.

We conclude that for each distribution the conditional approach gave higher values for the loglikelihood and that seems to be more appropriate to estimate the fraction of missing data. Further we found that the difference between the estimation techniques shows a clear tendency to become stronger, the more heavy-tailed the distribution is.

3.2 Loss Frequency Distributions

As discussed earlier, the severity and frequency distributions of the operational losses are inter-related. The frequency distribution needs to be adjusted according to the 'information loss', as discussed in the previous section. The frequency distribution needs to be adjusted by the fraction of missing data as indicated by the last column in tables 6 to 9. This will account for the truncated data and the higher true frequency of the loss events.

In the following we will use the Poisson distribution to model the frequency of different loss type categories. This is in line with the suggestions of the new Basel capital accord. The adjustments are necessary according to the estimated fraction of missing data and can be done according to

$$\lambda^* = \frac{\lambda^{obs}}{1 - P(X < H)}$$

Type	$< 1Mio.$	$> 1Mio.$	Complete	Fraction $< 1Mio.$
Relationship	59	470	529	0.111
Human	91	503	594	0.153
Processes	5	190	185	0.026
Technology	9	51	60	0.150
External	28	184	212	0.132

Table 11: Actual Losses below and above Threshold of 1 Million USD for 1985-2001

Based on the results of the previous section this implies that based on the estimation method the effect of the adjustment will be greater if the conditional estimation technique is used. Further the frequency adjustment will be dependent on the chosen distribution and adjusted frequencies will be clearly higher if a heavy-tailed distribution is assumed for the losses.

Type	λ	λ_{Tr}	λ_{EM}
Relationship	27.647	27.921	27.923
Human	29.588	29.758	29.758
Processes	11.176	11.207	11.207
Technology	3.000	3.033	3.034
External	10.823	10.925	10.926

Table 12: Parameter for estimated Exponential distribution λ, with frequency adjustment for truncation λ_{Tr} and EM truncation λ_{EM} according to Exponential distribution

Type	λ	λ_{Tr}	λ_{EM}
Relationship	27.647	28.971	31.586
Human	29.588	32.095	45.228
Processes	11.176	11.495	11.839
Technology	3.000	3.189	3.750
External	10.823	11.603	14.517

Table 13: Parameter for estimated Poisson distribution λ, with frequency adjustment for truncation λ_{Tr} and EM truncation λ_{EM} according to Lognormal distribution

Both effects are illustrated by tables 12 and 13. Table 12 shows frequency adjustments for the exponential distribution where hardly any change in the frequency parameter will happen, while for the lognormal distribution based on the estimated fraction of missing data especially for the estimation technique using the EM algorithm substantial frequency adjustments have to be conducted, see table 13. As we will see in the next section this also affects the estimated operational Value-at-Risk.

4 VaR simulation of Operational Losses

In this section we will consider the effects of the chosen estimation techniques on operational Value-at-Risk figures for the considered data. Based on estimates for severity and loss distributions one can generate scenarios for the losses and estimate Value-at-Risk figures for different loss-type categories. Clearly the VaR is dependent on the chosen loss distribution and the chosen process for modeling the frequencies.

Type	Mean	$VaR_{0.95}$	$VaR_{0.99}$
$Relationship_1$	0.2803	0.4123	0.4760
	(0.0009)	(0.0022)	(0.0028)
$Relationship_{Tr}$	0.2829	0.4151	0.4804
	(0.0007)	(0.0023)	(0.0033)
$Relationship_{EM}$	0.2806	0.4116	0.4754
	(0.0011)	(0.0016)	(0.0037)
$Human_1$	0.5214	0.7567	0.8731
	(0.0010)	(0.0039)	(0.0073)
$Human_{Tr}$	0.5240	0.7593	0.8789
	(0.0014)	(0.0034)	(0.0060)
$Human_{EM}$	0.5217	0.7592	0.8730
	(0.0014	(0.0051)	(0.0048)
$Process_1$	0.4123	0.7241	0.8987
	(0.0013)	(0.0040)	(0.0096)
$Process_{Tr}$	0.4145	0.7308	0.9021
	(0.0014)	(0.0043)	(0.0125)
$Process_{EM}$	0.4122	0.7237	0.8936
	(0.0012)	(0.0031)	(0.0095)
$Technology_1$	0.0270	0.0691	0.0960
	(0.0002)	(0.0008)	(0.0013)
$Technology_{Tr}$	0.0274	0.0706	0.0975
	(0.0002)	(0.0010)	(0.0012)
$Technology_{EM}$	0.0271	0.0698	0.0965
	(0.0002)	(0.0007)	(0.0011)
$External_1$	0.1157	0.2055	0.2537
	(0.0004)	(0.0013)	(0.0027)
$External_{Tr}$	0.1168	0.2070	0.2550
	(0.0004)	(0.0019)	(0.0029)
$External_{EM}$	0.1158	0.2052	0.2533
	(0.0004)	(0.0011)	(0.0035)

Table 14: Simulated Losses with Exponential distribution (10^{10}) - Threshold of 1 Million USD

The Basel accord suggests the use of a Compound Poisson process to aggregate the operational losses. The capital charge is based on the upper quantile of such aggregated distribution.

Let $\{N(t),\ t \geq 0\}$ be a homogeneous Poisson counting process and $\{X_j,\ j \geq 0\}$ be a sequence of iid random variables independent of $\{N(t),\ t \geq 0\}$. A compound process

$$S_N(t) = \sum_{j=0}^{N(t)} X_j, \quad t \geq 0 \tag{4.1}$$

is called a *compound Poisson process.*

The cumulative distribution of the compound Poisson process $S_N(t)$ is

$$P(S_N \leq x) = \begin{cases} \sum_{n=1}^{\infty} P(N(t)=n)\ F^{n*}(x) & x > 0 \\ P(N(t)=0) & x = 0 \end{cases} \tag{4.2}$$

where F^{n*} denotes the n-fold convolution with itself. In the following we will investigate the effects of the chosen distribution for severities and the approach for severity estimating and frequency adjustments to the loss data on operational VaR.

We will now use the results from sections 3.1 and 3.2 to analyze the impact of the 'information loss' due to the missing data, on the operational Value-at-Risk, and, hence, on the capital charge.

We consider three possible approaches a financial institution may choose th deal with the issue:

1. (Naive approach) Use the observed frequency $\hat{\lambda}^{obs}$ and fit the unconditional distribution to the truncated data. This oversimplified and misspecified approach will lead to biased estimates for the parameters of the loss distribution what is shown in [3].

2. Fit the unconditional distribution to the truncated data. Based on the estimated fraction below the threshold adjust the parameter $\hat{\lambda}^{obs}$ according to:

   $$\lambda^{Tr} = \frac{\lambda^{obs}}{1 - P(X < H)^{Tr}}$$

   This approach somehow considers that the missing data and threshold by adjusting the frequency, however does not account for the bias in the unconditionally estimated loss distribution. Since the expectation of the severity distribution will be overestimated and the frequency underestimated it will be the question which effect is stronger compared to using the estimates based on the EM algorithm.

Type	Mean	$VaR_{0.95}$	$VaR_{0.99}$
$Relationship_1$	0.2799	0.4620	0.5585
	(0.0013)	(0.0037)	(0.0055)
$Relationship_{Tr}$	0.3216	0.5146	0.6170
	(0.0009)	(0.0023)	(0.0051)
$Relationship_{EM}$	0.2808	0.4804	0.5893
	(0.0009)	(0.0023)	(0.0069)
$Human_1$	0.5206	0.8876	1.0935
	(0.0021)	(0.0050)	(0.0093)
$Human_{Tr}$	0.6303	1.0290	1.2449
	(0.0024)	(0.0054)	(0.0134)
$Human_{EM}$	0.5218	0.9025	1.1179
	(0.0017)	(0.0053)	(0.0075)
$Process_1$	0.4129	0.8670	1.1417
	(0.0029)	(0.0053)	(0.0166)
$Process_{Tr}$	0.4561	0.9280	1.2122
	(0.0031)	(0.0086)	(0.0207)
$Process_{EM}$	0.4144	0.9318	1.2711
	(0.0025)	(0.0062)	(0.0140)
$Technology_1$	0.0270	0.0827	0.1238
	(0.0003	(0.0012)	(0.0020)
$Technology_{Tr}$	0.0303	0.0891	0.1303
	(0.0002	(0.0008	(0.0010
$Technology_{EM}$	0.0269	0.0895	0.1415
	(0.0002)	(0.0011)	(0.0028)
$External_1$	0.1156	0.2471	0.3263
	(0.0006)	(0.0020)	(0.0043)
$External_{Tr}$	0.1373	0.2793	0.3635
	(0.0005)	(0.0019)	(0.0050)
$External_{EM}$	0.1166	0.2567	0.3477
	(0.0006)	(0.0030)	(0.0051)

Table 15: Simulated Losses with Gamma distribution (10^{10})- Threshold of 1 Million USD

3. Determine the MLE-estimates for the loss distribution with the presented EM-algorithm and determine the adjusted frequency by

$$\lambda^{EM} = \frac{\lambda^{obs}}{1 - P(X < H)^{EM}}$$

For simulations, draw losses from the unconditional distribution using the complete-data estimated parameters via the EM-algorithm, and use the complete-data frequency parameter. [3] show that this will lead to asymptotically correct results for the VaR.

The simulation results for the considered distributions are displayed in tables 14 to 17.

Dependent on the chosen the severity distribution the results are unambiguous. The more heavy-tailed the distribution the higher is the simulated Value-at-Risk. This can be observed for all loss type categories and is not surprising since the estimated heavy-tailed distributions provide higher risk figures in the tail.

But from the perspective of the impact of the including or ignoring the threshold in the parameter estimation procedure we can observe the following effect: for the light-tailed exponential distribution ignoring the threshold of 1 Million USD doesn't have too much impact on the simulated mean and Value-at-Risk. Using the EM algorithm for parameter estimation gives only small changes of the parameter λ. Therefore, also the frequency adjustment is so small that the effect on VaR figures could nearly be neglected. An interesting result is that for the light-tailed exponential for some loss categories the VaR is higher choosing the EM algorithm as estimation procedure while for some loss categories we get lower Value-at-Risk figures. However, in all event-type categories the changes in mean and VaR between the chosen methods are always lower than 5%. However, also recall that the exponential distribution couldn't provide a very good fit to the data.

For the heavy-tailed distributions we get a clear difference in VaR figures dependent on the chosen method. For most loss type categories the VaR determined by the approach using EM algorithm is between two and three times higher than for the biased method ignoring the threshold. This is especially true for the lognormal distribution. Further, we find that especially for the lognormal distribution the effect of frequency adjustment is strong enough to provide the highest VaR figures for the method using EM algorithm to determine the complete severity distribution and then adjusting the frequency. Thus, we recommend banks to include thresholds both in the severity and frequency estimation procedure to determine realistic operational VaR. However, further research and analytical calculations should be conducted.

5 Summary and Conclusive Remarks

This paper has demonstrated how dealing with the missing data under a prespecified threshold. Dealing with operational losses that are subject to reporting thresholds the approach showed that ignoring or including the threshold into the estimation procedure can seriously effect the estimated parameters of the loss distribution. If a frequency adjustment is conducted there are also great deviations between estimated frequencies based on the unconditional or conditional fit of the distribution. we also showed that the estimation technique also affects the estimated operational VaR and the operational risk capital charge. While for thin-tailed distributions the differences are rather small for more realistic distributions like the Lognormal or Weibull distribution ignoring the threshold may lead to a clear underestimation of operational VaR. Moreover, the empirical

Type	Mean	$VaR_{0.95}$	$VaR_{0.99}$
$Relationship_1$	0.2221	0.3931	0.5002
	(0.0012)	(0.0020)	(0.0050)
$Relationship_{Tr}$	0.2507	0.4330	0.5434
	(0.0008)	(0.0029)	(0.0026)
$Relationship_{EM}$	0.2509	0.5368	0.8019
	(0.0007)	(0.0027)	(0.0123)
$Human_1$	0.2802	0.5329	0.7117
	(0.0011)	(0.0029)	(0.0108)
$Human_{Tr}$	0.3299	0.6029	0.7933
	(0.0012)	(0.0063)	(0.0143)
$Human_{EM}$	0.4210	1.1015	1.9853
	(0.0037)	(0.0187)	(0.0693)
$Technology_1$	0.0305	0.1028	0.1705
	(0.0005)	(0.0019)	(0.0047)
$Technology_{Tr}$	0.0333	0.1083	0.1787
	(0.0004)	(0.0014)	(0.0052)
$Technology_{EM}$	0.0327	0.1288	0.2749
	(0.0005)	(0.0026)	(0.0120)
$External_1$	0.0860	0.2048	0.3033
	(0.0006)	(0.0013)	(0.0052)
$External_{Tr}$	0.1012	0.2297	0.3311
	(0.0006)	(0.0022)	(0.0052)
$External_{EM}$	0.1014	0.3107	0.6023
	(0.0008)	(0.0078)	(0.0259)

Table 16: Simulated Losses with Weibull distribution (10^{10})- Threshold of 1 Million USD

analysis of this paper suggests that the impact is more severe for more heavy-tailed distributions. However, more research on this issue has to be conducted.

We summarize other important remarks:

Other adjustments to frequency and severity distributions, and their aggregation, can be considered, in addition to those presented in the empirical analysis, other than the EM algorithm. Some banks, for example, may fit conditional rather than unconditional distribution to the losses, given that they exceed a given threshold. We do believe, however, that the EM algorithm provides the closest approximation to the true parameters of the assumed distribution, given that no information is provided on the number and magnitudes of the missing data other than their upper and lower bounds.

The λ parameter in the frequency distribution is assumed to be constant regardless of the loss magnitudes associated with it, in this paper. This can be of course generalized and a varying frequency parameter can be considered instead. One possibility would be to use the Cox processes.

This paper considered the simplest compound process for the aggregated losses,

Type	Mean	$VaR_{0.95}$	$VaR_{0.99}$
$Relationship_1$	0.2354	0.5289	0.9236
	(0.0014)	(0.0023)	(0.0055)
$Relationship_{Tr}$	0.2484	0.5497	0.9593
	(0.0009)	(0.0031)	(0.0029)
$Relationship_{EM}$	0.3415	0.8892	1.9364
	(0.0008)	(0.0029)	(0.0128)
$Human_1$	0.3343	0.8159	1.6163
	0.0066	0.0166	0.0823
$Human_{Tr}$	0.3598	0.8725	1.7142
	0.0055	0.0202	0.0716
$Human_{EM}$	0.9285	2.8411	9.2153
	0.0436	0.1375	0.7673
$Process_1$	0.5742	1.8005	4.5553
	0.0103	0.0382	0.1584
$Process_{Tr}$	0.5891	1.8138	4.4968
	0.0162	0.0554	0.2127
$Process_{EM}$	0.7936	2.5923	7.3580
	0.0267	0.0780	0.4281
$Technology_1$	0.0466	0.1732	0.5205
	0.0015	0.0041	0.0263
$Technology_{Tr}$	0.0490	0.1813	0.5220
	0.0013	0.0044	0.0303
$Technology_{EM}$	0.0949	0.3055	1.2925
	0.0070	0.0124	0.0806
$External_1$	0.0861	0.2457	0.5277
	0.0010	0.0053	0.0217
$External_{Tr}$	0.0922	0.2626	0.5685
	0.0013	0.0038	0.0341
$External_{EM}$	0.1706	0.5451	1.6829
	0.0101	0.0184	0.1316

Table 17: Simulated Losses with Lognormal distribution - Threshold of 1 Million USD

with the Poisson frequency distribution. However, we do not rule out the possibility of using other forms of compound processes. Some evidence [4] suggests that the inter-arrival times do not follow the Exponential distribution. These issues are also left to future work.

Acknowledgements

The authors are thankful to S. Stoyanov from the Finanalytica Group for computational assistance. A. Chernobai is grateful to the hospitality of the Institute for Statistics and Mathematical Finance, University of Karlsruhe, where the pa-

per was written during her visit in October-December 2004.

References

[1] BIS. Consultative document: operational risk. `www.bis.org`, 2001.

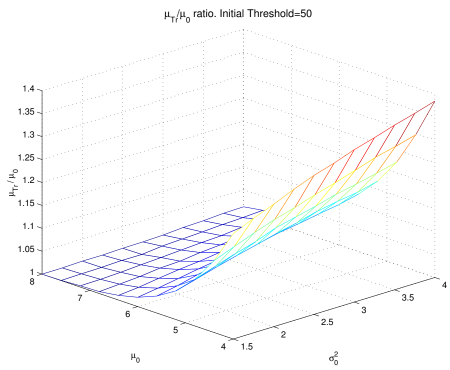
[2] BIS. The 2002 loss data collection exercise for operational risk: summary of the data collected. `www.bis.org`, 2003.

[3] A. S. Chernobai, C. Menn, S. Trück, and S. Rachev. A note on the estimation of the frequency and severity distribution of operational losses. *The Mathematical Scientist*, 30(2), 2005.

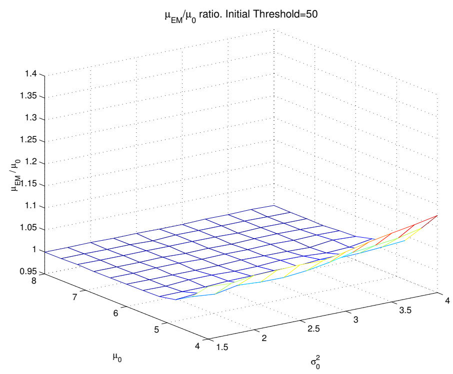
[4] A. S. Chernobai and S. Rachev. Stable modelling of operational risk. In M. G. Cruz, editor, *Operational Risk Modelling and Analysis. Theory and Practice*, pages 139–169, London, 2004. Risk Books.

[5] M. G. Cruz. *Modeling, Measuring and Hedging Operational Risk*. John Wiley & Sons, New York, Chichester, 2002.

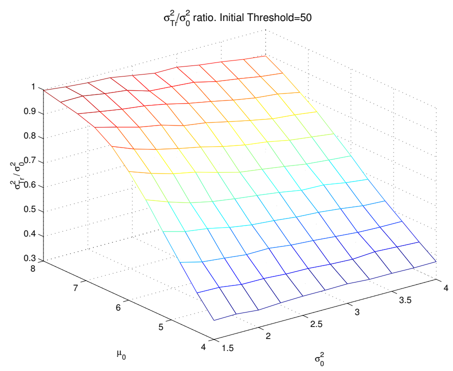
[6] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)*, 39(1):1–38, 1977.



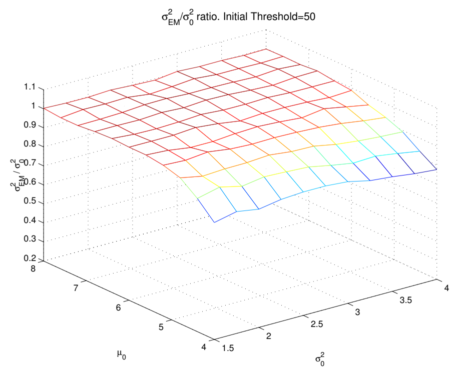
(a) μ_{Tr}/μ_0



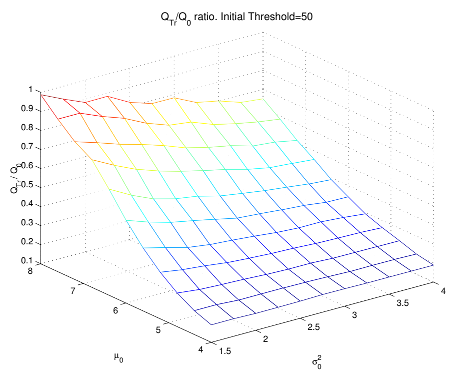
(b) μ_{EM}/μ_0



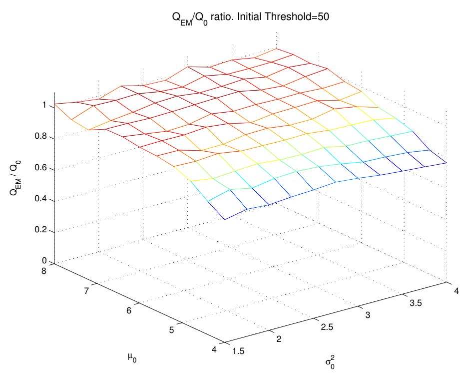
(c) σ^2_{Tr}/σ^2_0



(d) σ^2_{EM}/σ^2_0



(e) Q_{Tr}/Q_0



(f) Q_{EM}/Q_0

Figure 3: Robustness check for Lognormal distribution. Threshold = 50. A sample of 100,000 points is above each cutoff level.