

Challenges in Measuring Operational Risk from Loss Data

Eric W. Cope
IBM Zurich Research Lab
erc@zurich.ibm.com

Giulio Mignola
Intesa Sanpaolo S.p.A.
giulio.mignola@intesasanpaolo.com

Gianluca Antonini
IBM Zurich Research Lab
gan@zurich.ibm.com

Roberto Ugoccioni
Intesa Sanpaolo S.p.A.
roberto.ugoccioni@intesasanpaolo.com

September 2, 2009

Abstract

Under the Advanced Measurement Approach of the Basel II Accord, banks are required to measure their total annual operational risk exposures at the 99.9th percentile of the loss distribution. We examine the possibility of meeting this measurement standard given the amount of operational loss data that is currently available from either internal or external sources. We also survey some difficulties that arise in applying the Loss Distribution Approach to computing operational risk exposures, as well as in validating the capital models. Finding many of these problems insurmountable, we suggest some changes to the regulatory framework that would circumvent these difficulties.

The views expressed in this article are the authors' own and do not necessarily reflect the viewpoints of IBM, Banca IntesaSanPaolo, or the ORX consortium.

1 Introduction: The Regulatory Challenge

In the wake of the corporate scandals of the early 2000's and several high-profile bank losses, operational risk – which encompasses such diverse risks as criminal activity, legal liability, process execution errors, disasters, and systems failures – came to the foreground as a major source of impact to financial institutions. With the publication of the Basel II Accord in June 2004, operational risk assumed a position of similar importance alongside market and credit risk in the regulation of the banking industry. One of the three pillars of the Basel II framework is that banks set aside enough “regulatory capital” to self-insure against rare but severe operational losses. Three approaches to determining regulatory capital were set out; two of which follow simple formulas based on gross income, but the third allows qualifying banks to measure risks based instead on an internally developed empirical model. In recognition of the lack of standard models and data for measuring operational risk, this so-called Advanced Measurement Approach (AMA) provides banks with considerable freedom in the development of these internal models.

However, the measurement standard set by the AMA is extremely high. To quote the US Final Rule on the implementation of Basel II standard,

A bank using the AMA must demonstrate to the satisfaction of its primary Federal supervisor that its systems for managing and measuring operational risk meet established standards, including producing an estimate of operational risk exposure that meets a *one-year, 99.9th percentile soundness standard*.¹ (Emphasis added.)

The 99.9th percentile standard implies that banks estimate the magnitude of an annual loss that would occur on average every thousand years. Under the AMA, banks should derive their regulatory capital charges by reasoning from a variety of quantitative and qualitative factors, including internal data, external data, scenario analyses, and business environment and internal control factors. In this paper, we shall focus on the challenges of estimating the 99.9th percentile of the distribution of total annual operational losses based on internal and external loss data.

The line of argument in this paper may be summarized as follows. We first review some relevant properties of operational loss distributions that make estimation of high percentiles especially challenging (Section 2). Next, we show that the amount of data available for estimating high percentiles is insufficient, even if external data is taken into account (Section 3), and that extrapolation beyond the observed data is necessary. In Section 4, we argue that extrapolating far beyond the observed data cannot be justified in the context of extreme operational losses. Special attention is paid to the suitability of Extreme Value Theory (EVT) as a potential justification for extrapolation. Next, we demonstrate that the most common approach to

¹Federal Register, Vol. 72, no. 235, p. 69294.

estimating capital, the Loss Distribution Approach (LDA), can be overly sensitive to a small set of low-frequency, high-severity losses, further increasing the data sufficiency requirements and increasing the volatility and uncertainty of the capital estimates (Section 5). We discuss the requirements and possibilities for validating capital models in Section 6. Finally, we summarize in Section 7 the challenges in measuring operational risk and provide some suggestions as to how banks can better meet these challenges, as well as how the regulatory standard can be modified to create a more accessible and fair risk measurement standard.

2 The nature of operational loss distributions

Operational losses are commonly characterized as having a “heavy-tailed” distribution. In mathematical probability, a heavy-tailed distribution is one whose tails are not exponentially bounded (Asmussen, 2003). Of course, as exponential boundedness is an asymptotic property of loss distributions, we cannot practically verify based on limited loss data if the true distribution of operational losses really meets this criterion. Rather, when we speak about “heavy-tailed data,” we typically mean that even in a data set of moderate size, we observe losses ranging over several orders of magnitude, such that the largest losses may be hundreds, thousands, or even tens of thousands times greater than the median value. Typically, heavy-tailed distributions such as the lognormal, Pareto, or other power-law type distribution provide a good in-sample fit to operational loss data sets. A rule of thumb is that for every additional “9” added to the decimal, the quantile level increases by a multiplicative factor, i.e., the 99th, 99.9th, and 99.99th percentile represent different orders of magnitude. By contrast, for light-tailed (exponentially bounded) distributions, the difference between these quantile levels is within a constant value. See Figure 1 to view the range of values covered by a typical heavy-tail distribution.

Heavy-tailedness can be observed in nearly every category of operational losses, regardless of the event type or the business line. This creates some difficulties in the estimation of properties of the underlying distribution, because individual losses can be so large that they dominate the effects of all other losses in a sample. We list below a few of the consequences of working with heavy-tailed loss data:

- *Instability of estimates.* A single, extreme loss can cause drastic changes in the estimate of the mean or variance of the distribution, even if a large amount of loss data has already been collected. For example, suppose we observe 100 losses generated from a Lognormal(10,2) distribution with sample mean 162,750, which is a distribution that is representative of the losses we observe in many units of measure. If a once-per-thousand loss event (99.9th percentile) is observed, the estimated mean value will increase to 266,525 – a 64% increase in the mean as a result of a single loss value alone. Estimates of sample averages of

Extreme Values of a Heavy-Tailed Distribution

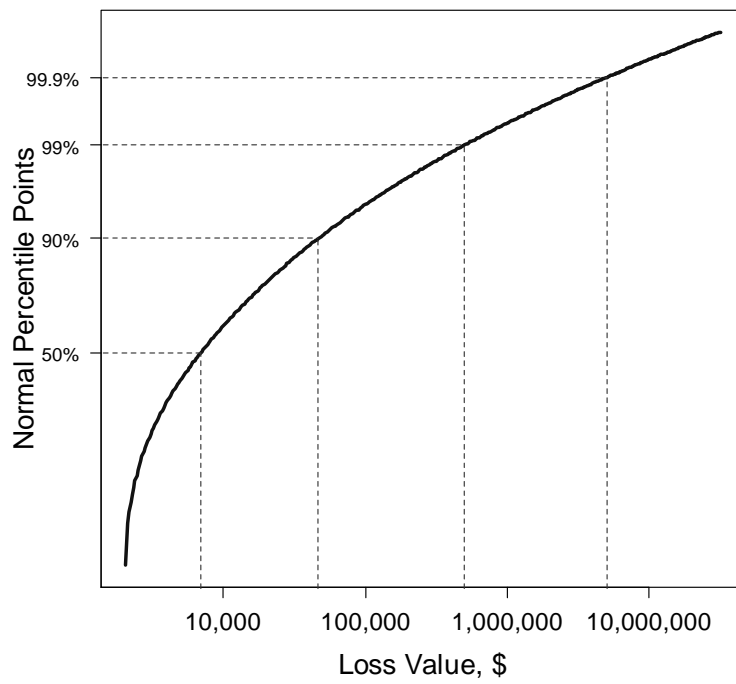


Figure 1: Probability plot of a heavy-tailed distribution.

functions of heavy-tailed data can therefore be highly volatile.

- *Dominance of sums.* Annual total operational losses for a unit of measure will typically be driven by the value of the most extreme losses. For example, out of a set of 6,593 operational losses recorded in the Operational Riskdata eXchange (ORX) database (ORX, 2009) in the Retail Banking business line for the Clients, Products, and Business Practices event type, the ten largest losses account for 44% of the total losses, and the 100 largest losses account for nearly 75% of the total loss value. In addition, the following theoretical result shows that for heavy-tailed loss distributions, high quantiles of the total annual loss distribution are primarily determined by high quantiles of the distribution of the maximum loss: if S_n represents the sum of n heavy-tailed losses, and M_n represents the maximum of these n losses, then for large values of x ,

$$P(S_n > x) \approx P(M_n > x). \quad (1)$$

In fact, for very heavy-tailed loss distributions having infinite mean, the total losses S_n will remain within a constant multiple of the maximum loss M_n , even for very large amounts of loss data (Embrechts et al., 1997). This indicates that a single loss can be responsible for most of the total losses experienced within a unit of measure in a year.

- *Dominance of mixtures.* If losses that are regularly generated in two different units of measure are pooled together, with the distribution of one unit of measure having a heavier tail than the other, then the tail of the distribution of the total losses will typically follow the tail of the distribution of the heavier-tailed loss category. This follows from the earlier statements regarding the dominance of sums, since for large enough data sets, the largest values within the pooled data set will come from the heavier-tailed unit of measure with very high probability. These largest values will again primarily drive the value of the total losses.

Each of these facts is significant to the estimation of annual total operational losses for a bank, because the capital calculations require that one combine the effects of losses from many units of measure where loss distributions may vary greatly in severity. As we shall discuss, these properties imply that both the value of capital estimates, and the amount of uncertainty around those estimates, will be primarily driven by single, extreme loss events.

3 Data sufficiency

To begin to understand how much data is needed to estimate the high percentiles of a loss distribution, let us consider an approximation for the standard error of a nonparametric estimator of the quantile q_p at probability level p , based on a large number n of samples drawn from a heavy-tail distribution with density function f :

$$\frac{1}{f(q_p)} \sqrt{\frac{p(1-p)}{n}} \quad (2)$$

(Kiefer, 1967). Note the presence of the density function in the denominator, which will be very small for values of p close to one, which has the effect of increasing the estimation error. We will therefore need approximately $400p(1-p)/(f(q_p)q_p)^2$ samples or more to get a relative error of less than 10 percent (where relative error is measured as two standard errors divided by the true quantile value). For example, suppose the distribution function of losses is of the Pareto form

$$F(x) = 1 - x^{-\alpha} \quad \alpha > 0, x \geq 1. \quad (3)$$

Then the required number of samples to achieve less than $\pm 10\%$ relative error is 277,500 when $p = 0.999$ and $\alpha = 1.2$. Compare this to an Exponential distribution with rate 1, where we need around 8,400 data points to estimate the 99.9th percentile to within 10% relative error. It is significantly more difficult to get accurate quantile estimates for heavy-tailed distributions.

For a more direct estimate of the data requirements, consider Table 1, which lists the widths of a 95% confidence interval for the 99.9th percentile of a Lognormal(10,2) distribution, as a function of the number of data points collected. The true value of the 99.9th percentile of this distribution is 10.6 Million. If we have 1000 data points to estimate this quantity, we see that 19 times out of 20, our estimates fall within a range that is as wide as 20 million, or a range of uncertainty that is twice as large as the value we are estimating. As we collect more data -- orders of magnitude more -- it still takes a long time to reduce the range of uncertainty to a reasonable error. Even with 100,000 data points, our estimates are still not within $\pm 10\%$ of the true value, and it requires nearly a million samples to reduce the uncertainty to $\pm 5\%$. This is nearly ten times more loss data than has been collected in the ORX consortium to date, whose database is the largest commercial database and includes those losses in excess of €20,000 from more than 50 banks over a seven-year period.

Clearly, we need to collect a large amount of data on individual losses to compute high quantiles of the severity distribution. However, the regulations do not call for the 99.9th percentile of the single-loss severity distribution to be measured, but rather the 99.9th percentile of the annual total loss distribution. To directly estimate this quantity, we would need to collect

Number of data points	Width of a 95% confidence interval for the 99.9th percentile (Millions)
1,000	20.0
10,000	7.7
100,000	2.5
1,000,000	0.8

Table 1: Widths of confidence intervals for estimates of the 99.9th percentile of a Lognormal(10,2) distribution in millions, as a function of the amount of data collected.

many thousands of years of loss data. However, since the regulations came in force, most banks have systematically collected data over a period ranging between five and ten years. An estimate of the once-per-thousand-year loss event from this data would require a 100- to 200-fold extrapolation beyond the observed data, which is clearly infeasible.

Yet how can we estimate this figure if there is so little internal data? One might wonder, since we have much more data on individual losses than on total annual losses, if it would be possible to achieve an improved estimator of the total annual loss distribution based on an estimate of the single-loss severity distribution. Indeed, the relationship (1) provides a clue as to how this can be done. Suppose F denotes the single-loss distribution, and $F^{(n^*)}$ the distribution of the sum S_n of n losses (i.e., the n -fold convolution of F), and let \bar{F} denote the complement $1 - F$. For a class of heavy-tailed losses, it can be shown (Embrechts et al., 1997) that the probability that the maximum loss M_n exceeds a level x , $P(M_n > x)$ is approximately equal to $n\bar{F}(x)$ for large x , which implies from (1) that

$$\overline{F^{(n^*)}}(x) \approx n\bar{F}(x).$$

This in turn implies that an approximation for a high quantile of the distribution of total annual losses can be expressed in terms of the single-loss distribution:

$$(F^{(n^*)})^{-1}(p) \approx F^{-1}(1 - (1 - p)/n), \quad (4)$$

where F^{-1} denotes the inverse cdf or quantile function (Böcker and Klüppelberg, 2005). We shall refer to equation (4) as the *single-loss approximation*.

Unfortunately, even though the single-loss approximation allows us to use estimates of loss severities to compute annual total loss distributions, it does not reduce the data required to achieve a given level of accuracy. To see this, suppose a bank experiences an average of 100 losses per year in a unit of measure. The single-loss approximation says that an estimation of the 99.9th percentile of the distribution of annual losses is approximately equal to the 99.999th percentile of the single-loss approximation, or the size of the once-per-100,000 loss event. To estimate this quantity without extrapolation, the bank would need to collect over 100,000 samples – which would require 1,000 years of data at the current rate of loss occurrence. This is exactly the minimum amount of data needed to estimate the 99.9th percentile of the annual loss distribution without extrapolation. Heuristically, since the once-per-thousand year annual loss will be approximately equal to the maximal loss incurred in that year, and the maximal loss that year will approximately be the largest loss experienced in 1,000 years, we need to compute the maximal loss among 100,000 events. This indicates that there is “no free lunch” in using the single loss distribution to estimate annual losses, and therefore must extrapolate well beyond

the available data to estimate capital.

4 Extrapolating beyond the data

Since extrapolation beyond the data is inevitable, we next examine the possible grounds that may be given for justifying a given extrapolation, and some of the ramifications involved in selecting a model. We first note that some parametric model for the loss distribution must be applied in any extrapolation, since nonparametric estimation does not provide any guidance outside of the range of observed data. How a parametric model performs in extrapolation depends on how well it corresponds to the underlying loss-generating mechanisms. There are several options for how to model a data set containing extreme values (cf. Lindgren and Rootzén (1987)):

1. A single mechanism is responsible for all observed losses, and this mechanism can be assumed to produce any future losses that exceed the levels currently seen in the data;
2. Multiple mechanisms produce different loss events, some of which may be more likely to produce extreme events;
3. The most extreme values are anomalous and/or do not follow a continuous pattern established by the decay in probabilities observed in the rest of the data.

From a statistical point of view, we should hope that we are in situation (1). In this case, we can apply a single parametric model to all of the observed losses and using this distribution to extrapolate. Ideally, such purely statistical reasoning would be grounded in a deeper understanding of how losses are generated, and such that the parametric form reflects an understanding of how the likelihood of such events decays with large loss values. This situation is one that is more frequently encountered in engineering and the physical sciences, where extreme-value models are in common practice in areas such as hydrology and reliability testing (Castillo et al., 2004; de Haan, 1990; de Haan and de Ronde, 1998).

However, there are many reasons to assume that situations (2) or (3) may be the case when dealing with operational losses. First of all, according to the practice established in the Basel Accord, losses are categorized according to seven primary loss event types, which include such diverse events as internal and external fraud, employment practice and safety issues, system failures, process execution errors, disasters, regulatory penalties and legal liabilities. At this level of granularity of the loss categories, it is inevitable that several loss-generating mechanisms will be at work within any given loss event category. Some of these mechanisms will be associated with losses of higher severity than others. For example, in the Asset Management business line, a series of high severity events in the Clients, Products, and Business Practices event type

occurred in 2003 due to fines for illegal trades emerging from late trading; these losses are much larger than any other losses in this category (Braceras, 2004). If other large losses are likely to arise in the future as a result of similar events, then a model of extreme events should not be estimated based on the other, less severe events in this category that are unrelated to market timing issues, otherwise the results may be subject to bias.

The effects of such data “contamination” were described in a previous study (Nešlehová et al., 2006), which found that a few contaminating data values could have a large effect on the resulting extrapolation. We can confirm this on the basis of ORX data, where some of the largest losses can be found in Clients, Products, and Business Practices in the Corporate Finance category. From January 2002 through June 2008, 283 losses were reported in this category, of which two exceed €2 billion. In Figure 2 we show the results of fitting a lognormal distribution to the entire data set, as well as to the same data set excluding the three largest observations. The estimates were obtained by using maximum likelihood techniques to fit a lognormal distribution, adjusted for the fact that ORX data is subject to a reporting threshold and therefore the data are truncated from below at the €20,000 level. In general, we can see from the figure that the distribution fits the data quite well, although they begin to diverge from each other at high quantile levels. When we fit the lognormal distribution to the entire dataset using maximum likelihood, the 99.9th percentile is around €5 billion. When we remove the largest three values from the data set, the quantile estimate drops by 65%, to €1.8 billion. Therefore, single loss events can have a major effect on quantile estimates.

Another point to consider is how quickly the loss-generating mechanisms change with time, so that data becomes less relevant as time passes. With rapid pace of changes in recent years in regulatory, legal, and political environments, increases in the sophistication of criminal activities and fraud, changes in the technological landscape and innovation in financial products. Furthermore, as banks pay more attention to governance, risk, and compliance activities, they are able to exert better control over loss events through prevention and mitigation actions. Banks of course also change over time. They grow in size, merge or divest business units, enter new markets, and restructure their organizations. Finally, the occurrence of an extreme loss at a bank will usually bring about large-scale changes in the way the bank conducts and controls its business, such that the occurrence of a similar loss of equal magnitude arising again is greatly diminished. In such cases, it is probably best to discard these data values, as in situation (3), but we are then in the unfortunate position of throwing out the only data that is likely to be in the range of the losses we wish to predict.

Indeed, even if we do assume that we are in situation (1), there are many structural reasons why it may be inadvisable to extrapolate from the observed data to unobserved regions of the loss distribution. First, there may be hard limits to the amount that can be lost in a

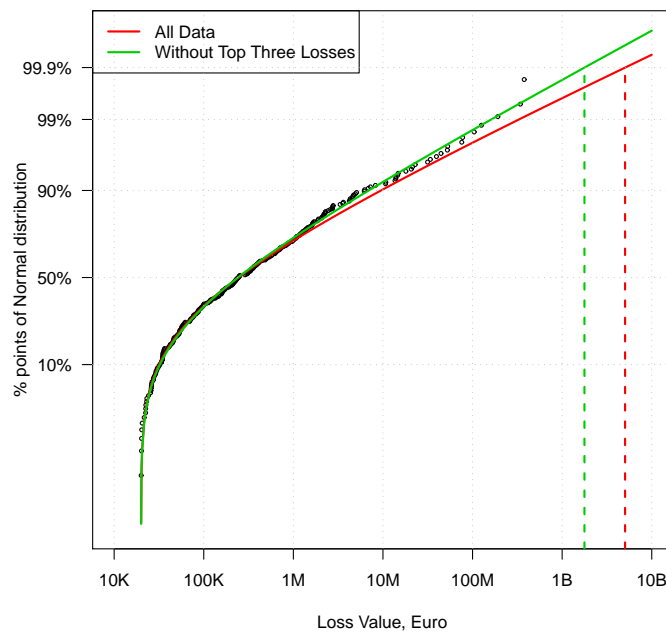


Figure 2: Lognormal fits to ORX Corporate Finance / Clients, Products, and Business Practices loss data, using all data and excluding the largest loss and the two largest losses. We have omitted plotting the three largest losses in this category for reasons of confidentiality. The data are plotted based on the quantile levels of the data set with the top three values removed.

Data set	Fitted 99.9th percentile (MM)
All data	5,030
Excluding three largest losses	1,774

Table 2: 99.9th percentile estimates for lognormal fits to ORX Corporate Finance / Clients, Products, and Business Practices loss data.

given event. For example, underwriting risks account for some of the largest operational losses on record, such as the massive losses associated with the collapses of Enron and Worldcom. With such risks, the amount of loss cannot exceed the total underwritten amount. No bank should extrapolate beyond these hard limits in projecting the once-per-thousand-year event from observable incidents. These considerations underline the fact that operational losses in banks are highly diverse and complex events that probably require a much more detailed treatment than a simple extrapolation of a parametric fit to large groupings of data on loss events.

However — let us suppose that we really are in situation (1) above — and that we can reasonably model both the central and extreme data values using a simple statistical model. In fact, let us even assume that we know the parametric family of distributions from which the losses are generated. How accurate is this model if we use it to predict the likelihood of events far in the tail of the distribution? Suppose a bank experiences a set of N losses per year in a category, where the loss amount given an event is distributed as a Lognormal(10,2) random variable. Recall from earlier discussion that the 99.9th percentile of this distribution is 10.6 million; we shall use the single-loss approximation for the 99.9th percentile of the aggregate loss distribution in the cases where the losses per year $N = 10$ and $N = 100$. Table 3 shows the range of values for estimates based on different amounts of data, where the distribution parameters were fit using maximum likelihood. As in the previous Table 1, we see that the confidence intervals are extremely wide in relationship to the true quantile values, which are 37.4 and 111.5 million, respectively.

In the previous paragraph, we assumed that we knew the parametric form of the distribution from which the losses were generated. Of course in practice, one must select an appropriate distributional form, which may not be easy (Mignola and Ugocioni, 2006). The difficulty of making a selection is illustrated in Figure 3, which shows actual loss data (in Euro) from the ORX database from Clients, Products, and Business Practices in the Private Banking business line. We fitted both a lognormal and a Generalized Pareto distribution using maximum likelihood methods to 752 data values exceeding the median of the complete set of reported data in this loss category. We fitted truncated versions of the lognormal and Pareto, with a truncation point equal to the median value of €64,572. That is, if $F(x; \theta)$ is the distribution under parameter θ , then the truncated version F_t ($t = 64,572$) is

$$F_t(x; \theta) = 1 - \frac{1 - F(x; \theta)}{1 - F(t; \theta)}.$$

The Generalized Pareto distribution was parameterized according to

$$F(y; \mu, \sigma, \xi) = 1 - \left[1 + \xi \left(\frac{y - \mu}{\sigma} \right) \right]^{-1/\xi},$$

Number of Data Points	Width of 95% conf. int. for SLA	
	10 Losses p.A.	100 Losses p.A.
100	100 (10 yrs)	355 (1 yr)
1,000	27 (100 yrs)	90 (10 yrs)
10,000	8 (1,000 yrs)	28 (100 yrs)

Table 3: Widths (in millions) of 95% confidence intervals for out-of-sample estimates of high quantiles of a Lognormal(10,2) distribution, corresponding to the single-loss approximation for total annual losses based on a frequency of 10 and 100 loss events per year. The number of years of data required to gain this level of accuracy under the assumed loss frequency is listed in parentheses beside each entry.

where μ is a location parameter set equal to t . In the figure, we see that both distributions appear to fit the body of the distribution very well, but begin to diverge in the tail – although it is hard to determine which fits better simply by a visual inspection. Both fitted distributions pass all goodness of fit tests, including the Cramer-von Mises, Anderson-Darling, and right-tail Anderson-Darling, at a significance level of 0.05, with nearly equal p -values. We note that among the banks in the ORX consortium, the average annual frequency of losses in this category is 8, with a maximum frequency of 32. Table 4 displays the fitted parameter values and 95% confidence intervals for the single-loss approximation for a bank experiencing 20 losses in this category per year, and Table 5 shows the differences in the estimated quantile values between the two distributions. From this, we see that using the same data, two banks may arrive at capital estimates for this loss category that differ by billions of Euro, depending on the distribution that they decide to fit.

With regard to the choice of models for extrapolation, some statisticians would argue that, *ceteris paribus*, Extreme Value Theory (EVT) provides a clear reason for preferring a Generalized Pareto Distribution (GPD) when extrapolating beyond data. EVT is a branch of mathematical probability theory that addresses the asymptotic distributions of extreme events. Just as the Central Limit Theorem provides an asymptotic theory for the distributions of averages, EVT provides an analogous set of statements regarding maxima of data sets and the distribution of data values exceeding high threshold levels. A key theorem (Balkema-de Haan-Pickands, Panjer (2006)) states that if a distribution F obeys certain regularity criteria, then the distribution of values exceeding a threshold converges in shape to a GPD distribution as the threshold increases. We should note that the GPD is really a family of distributions that encompass a wide range of tail shapes that are determined by a shape parameter ξ . However, if we can reliably estimate the parameters of the GPD from the available data, and if we can both assume that the regularity conditions apply and ensure that we are dealing with losses that are “large enough” to be within the asymptotic regime of the theorem, then we have a valid means of extrapolating beyond our data set.

It is important to remember, however, that EVT is a mathematical theory, and not a scientific law, and that its application to operational loss data is appropriate only if there are empirical grounds for doing so. Of course, we must first assume that we are operating in scenario (1) above; otherwise, applying EVT or any other method of extrapolating from the data would be unjustified. Second, even if we assume that the regularity conditions do apply (essentially, they say that the distribution of excesses over a threshold, properly scaled, will approach some limiting distribution), we must still determine whether or not we are operating with a high enough threshold value for the limiting values to be appropriate. There is evidence that for some distributions that fit well to many data sets, the rate of the convergence of the excess

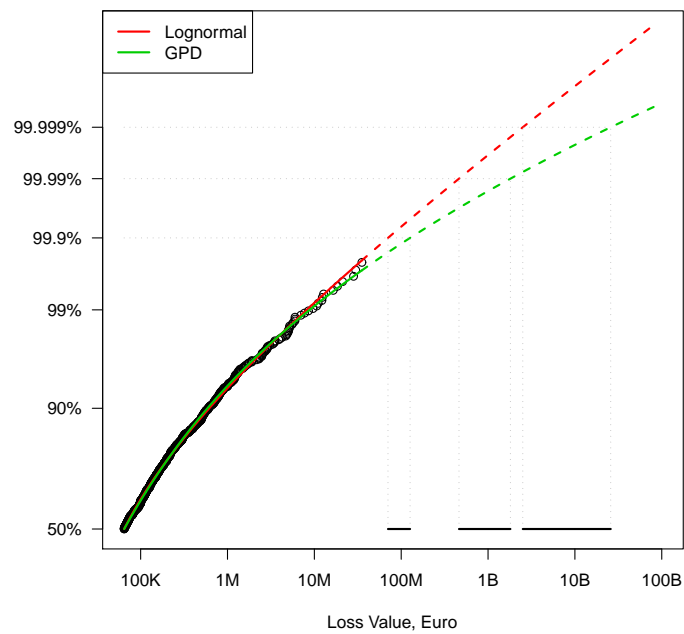


Figure 3: Lognormal and Generalized Pareto fits to ORX Private Banking / Clients, Products, and Business Practices loss data. We have omitted plotting the largest three loss values for reasons of confidentiality.

Distribution	Parameter	Estimate	Single-Loss Approximation (MM)	95% Confidence Interval for SLA
Lognormal	μ	4.598	784	(462, 1,610)
	σ	3.525		
GPD	σ	111,500	4,030	(2,220, 7,490)
	ξ	1.155		

Table 4: Maximum likelihood parameter estimates, single-loss approximation for a bank experiencing 20 losses per year, and 95% confidence intervals for the lognormal and Generalized Pareto (GPD) fits to the ORX Private Banking / Clients, Products, and Business Practices loss data. Confidence intervals were obtained through Monte Carlo simulation assuming that the distribution parameters followed a Normal distribution with asymptotic covariance matrix

Percentile	Diff. in Quantile Estimate (MM)
99.9	56
99.99	1,345
99.999	23,337

Table 5: Differences in fits at high percentiles using Lognormal and Pareto maximum likelihood estimates for the distribution of ORX Private Banking / Clients, Products, and Business Practices loss data.

distributions to the limiting distribution can be extremely slow. This was shown by Mignola and Ugoccioni (2005) for the lognormal distribution and by Degen et al. (2007) for the g-and-h distribution. These analyses indicate that quantile estimates based on EVT can be inaccurate, especially in extrapolation, as the amount of available data does not allow us to set a threshold nearly high enough to place us within range of the asymptotic limit. Therefore, the use of EVT should not rule out the fitting of other distributions, nor should any extrapolation be performed without attempting to understand the limitations of the model or the characteristics of the data-generating processes.

5 Loss Distribution Approach

In the previous section, we considered statistical estimation of severity models of losses within single units of measure only. However, in order to obtain an estimate of annual losses for the entire bank, the distributional estimates from each unit of measure must be combined in some manner. This is normally done using an actuarial method known as the Loss Distribution Approach (LDA). The LDA assumes that losses that occur within a unit of measure are independent and identically distributed (iid) according to some distribution F ; the annual frequency of losses N is distributed according to a given distribution, usually of the Poisson or Negative Binomial type. Annual losses for a unit of measure i may be expressed as the random sum

$$L_i = \sum_{j=1}^N L_{ij},$$

where the L_{ij} are iid losses according to distribution F_i , and are also independent of N . Typically, Fourier transform methods for computing distributions of sums of independent variables are employed to determine the distribution of L_i . Although the losses within a unit of measure are assumed to be independent, losses across units of measure are not. The total annual loss for a bank with k units of measure is then

$$L = \sum_{i=1}^k L_i.$$

To determine the distribution of L , a dependence structure (such as a copula) is estimated for the L_i 's, and a Monte Carlo simulation is typically performed to draw annual losses from each unit of measure with the correct dependence structure. Typically, there is a weak positive dependence across units of measure (Cope and Antonini, 2008).

Even if other approaches exist, the LDA framework is largely accepted as the standard framework for operational risk modeling. Although we believe that the main challenges related

to measuring operational risk are for the most part framework-independent, some of the assumptions of the LDA limit the accuracy of the models. For example, within the ORX data, we often observe trends and serial correlation in the time series of loss frequencies and severities among banks. Trends among a single bank’s loss categories may be attributable to a number of factors, including bank growth, mergers and acquisitions, improvements or changes in the loss reporting and the control environment, and the completeness of reporting among recently incurred losses. Other factors that may affect several banks simultaneously include changes in the legal and regulatory environment, and, importantly, fluctuations in the exchange rate.

A further aspect about temporal effects on operational losses is related to time dependencies between extreme losses, defined on the base of some in-sample quantile threshold, such as the 90th or 95th percentile. Looking at the distribution of inter-arrival times between these “spikes,” there is evidence of a departure from exponentiality, which one would expect if the LDA assumptions were correct (see Figure 4). Moreover, in some cases these spikes have the tendency to cluster, arriving within groups. One measure of this tendency is the *extremal index* of the process (see Embrechts et al. (1997) for a definition of the extremal index). Intuitively, such an index measures the reciprocal of the average size for clusters of large losses over time (e.g., a value of 1/2 indicates that, on average, large losses arrive in pairs). For iid processes, the extremal index is equal to 1. The estimated extremal index value associated with the spikes inter-arrival times shown in Figure 4 is equal to 0.202, confirming the evidence for dependencies over time, also among extreme loss values.

Issues around the applicability of the LDA have also been discussed in Frachot et al. (2004). Sources of uncertainty in the LDA calculation have also been discussed by Mignola and Ugocioni (2006). In the following sections, we shall focus on the accuracy of the results of the LDA model in the context of operational losses, given that the assumptions of the method are true.

5.1 Dominance of High-Severity, Low-Frequency Losses

The losses observed in various units of measure can be highly variable in terms of their frequency and severity. We mentioned earlier that Clients, Products, and Business Practices losses in the Corporate Finance business line are very rare, but can be highly severe, with the largest losses in the hundreds of millions, or billions. On the other hand, we also saw that External Fraud losses in Retail Banking alone make up around 30% of the total losses in the ORX database, and rarely exceed €1 million. A low-frequency, high-severity loss category contributes few losses, and so estimates of the annual loss distribution from that unit of measure are based on very few data points. Thus, not only are the high quantile estimates of the distribution extremely large, but also highly uncertain. As might be expected, the low-frequency, high-severity losses tend to dominate when one estimates high quantiles of the total annual loss distribution in the LDA

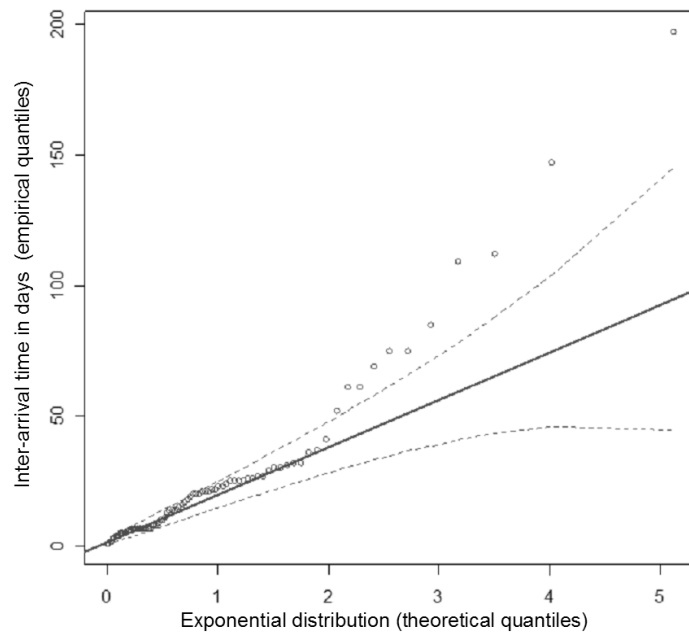


Figure 4: A quantile-quantile plot of interarrival times of exceedances of the 90th percentile for the Commercial Banking business line of one ORX institution. The spikes interarrival times are expressed in days, referring to the date of occurrence and considering only those losses occurred after 2002. The straight line indicates exponentiality and the dotted lines represent a 95% confidence level for a point-wise confidence envelope. At the high end, we see significant departures from exponentiality, indicating the clustering of extreme losses in time.

computation, not only in terms of the value of the estimate, but also in terms of the uncertainty around this estimate.

The effect of a low-frequency, high-severity loss category on the overall capital calculation can be seen in the following simulation experiment. Suppose that a bank experiences high-frequency, low-severity losses in ten separate business lines, each of which experiences losses in excess of €20,000 with severity distributed as a truncated Lognormal(8.00,2.34). This severity distribution is a good fit for losses in the ORX database in all event types excluding the Clients, Products, and Business Practices event type. The bank experiences 1,000 losses per year, with an average 100 losses per year in each of the business lines. It estimates the severity distribution for each business line from an external data set of 10,000 points. In addition, the bank has one additional unit of measure, corresponding to Corporate Finance in the Clients, Products, and Business Practices / Corporate Finance area. The bank experiences one loss per year in this unit of measure, with a severity distribution that is also truncated Lognormal with parameters (9.67,3.83), which describes the distribution of ORX data in this category well. The bank estimates this distribution based on approximately 300 data values from a consortium. Based on the error of estimating these distributions using maximum likelihood estimated from the available data, the median total bank capital from the ten business lines is €634 million Euro, with estimates ranging in a 95% confidence interval of width €98 million. The total bank capital with both the ten business lines and the additional unit of measure is €5,260 million, with estimates ranging in a 95% confidence interval of width €19 billion. We see that the one high-severity unit of measure, which accounts for 0.1% of the bank’s losses and 0.3% of the consortium data, contributes 88% to the total capital estimate, and 99% to the range of uncertainty of the estimates. Based on this realistic illustration, most of the regulatory capital in the bank and nearly all of the uncertainty in that estimate can be determined by only a small fraction of the data in one or two high-severity, low-frequency units of measure.

5.2 Negative Diversification Benefits

Another issue that arises when using quantiles of the loss distribution (often referred to as Value-at-Risk) to measure risk is that *negative* diversification benefits are often observed when the losses are heavy-tailed. In essence, this means that the total risk for the bank may exceed the sum of risks across the individual units of measure, a phenomenon that has been discussed in the operational risk literature (Nešlehová et al., 2006). This result is highly counterintuitive, as one normally thinks that a “diversified” portfolio of risky assets is less risky than a homogeneous portfolio. This effect can occur even when the losses in different units of measure are independent; in fact, the heaviness of the tail of the loss distributions has much more to do with the overall diversification benefit than does the correlation and dependence structure, at least

within the range that we observe correlations in loss data (Cope and Antonini, 2008). The effect is related to the concept of “incoherence” in risk measures (Arztner et al., 1999); Value-at-Risk is an incoherent risk measure and therefore can exhibit negative diversification benefits.

Letting $F_X^{-1}(0.999)$ denote the 99.9th percentile of the random variable X , we may define the *diversification benefit* at the τ quantile according to

$$D(\tau) = 1 - \frac{F_L^{-1}(\tau)}{\sum_{i=1}^k F_{L_i}^{-1}(\tau)}.$$

The degree of diversification benefit usually depends on the selected level of τ , and typically the heavier-tailed that the distributions F_{L_i} are, the smaller the diversification benefit will be. When losses are strictly positive, iid random variables whose distributions are regularly varying, it can be shown that as $\tau \uparrow 1$, $D(\tau)$ approaches a negative limit if and only if the random variables L_i have an infinite mean, which corresponds to a very heavy-tailed distribution (Embrechts et al., 1997). See Figure 5.

However, even for distributions that do have finite means, such as the lognormal, the diversification benefits can still be negative even at high levels of τ . For example, suppose the annual losses in ten units of measure are independently distributed as Lognormal(9,2). Based on a Monte Carlo simulation of 1 million samples, we estimate the diversification benefit $D(0.999) = 65.4\%$. However, if we replace one of these units of measure to have a loss distribution that is Lognormal(9,4), the estimated value of $D(0.999)$ drops to 1.2%. If we replace a second unit of measure in a similar way, then the overall diversification benefit becomes negative: $D(0.999) = -6.1\%$. What this implies with respect to low-frequency, high-severity distributions is that not only do they dominate the capital estimates, but the presence of one or more such distributions among the units of measure may further inflate the overall capital estimates by reducing the overall diversification benefit.

5.3 Sensitivity to Loss Categorization

To this point, we have been treating units of measure as exogenously determined groups of losses on which frequency and severity models are built. Units of measure may be organized according to business line and event type categories as described in the Basel II regulation. However, there is typically a very uneven distribution of losses across the business line-event type matrix. For example, in the ORX data around 30% of the reported losses fall into one cell category, External Fraud / Retail Banking, while 56 of the remaining 80 cell categories in the ORX classification include fewer than 0.5% of the overall losses. Even more sparsity is introduced if one considers additional classifications to further partition the set of losses, such as geographical region and

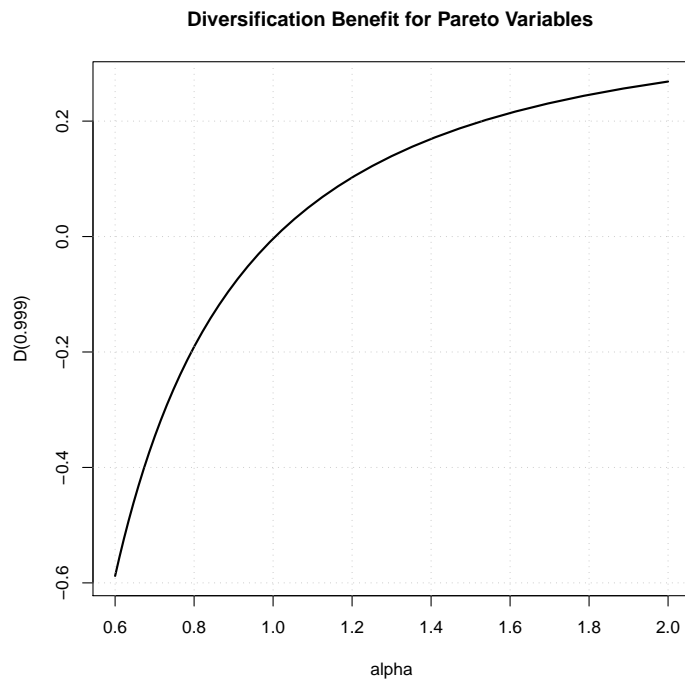


Figure 5: Diversification benefits $D(0.999)$ at the 99.9th percentile level for the sum of two independent loss distributions that are Pareto-distributed according to (3), for various values of the distributional parameter α . The range of α values displayed is typical for the tail behavior of operational loss distributions. For low values of α below 1, corresponding to heavier-tailed distributions, we observe negative diversification benefits.

firm size. In order to support an LDA analysis, however, sufficient data is required in each of these categories, and it is necessary to pool data in low-frequency loss categories that appear to follow a similar distribution.

However, the LDA can be sensitive to how the losses are pooled. For example, two units of measure that appear to have similar distributions, except that the most extreme losses occurred in only one of those units of measure, will typically give lower capital values when pooled than if they are modeled separately. To illustrate this, we simulated 200 data values from a Lognormal(10,2) distribution, and divided them equally between two units of measure to serve as sample data, except we always placed the largest three data values in the sample in the first unit of measure. We then estimated a lognormal distribution for the severity distribution of each unit of measure, as well as for the pooled data, and computed capital estimates using the LDA approach, assuming 10 losses per year in each unit of measure (20 for the pooled unit). We found that the capital estimates were nearly always higher in the case where the data were not pooled, with the average difference being about €4.6 Million, or about 9% of the average pooled capital estimate. This results from the fact that the unit of measure with the three largest data values will have a higher estimated severity, which again can have a dramatic effect on the tail behavior of the severity distributions. Note that there is a 25% chance that the largest three data values between two units of measure will occur in the same unit, so the simulation addresses a case that is not at all rare in pooling loss data.

In addition, the result of the LDA can also be sensitive to the method by which various units of measure are aggregated into the total. For a discussion, see Embrechts and Puccetti (2008).

6 Validating capital models

Recalling that a *model* is a simplified description of much more complex *reality*, it is essential that any model should be subject to a process of verification in order to prove that all the simplifying assumptions that are embedded in the model do not overly bias the results. Furthermore, the validation of those results is of fundamental importance if a regulatory value is attached to the model result, as we have in the case of operational risk capital models.

Considering the arguments put forward in this paper, the challenge to validate operational risk models is considerable. If one considers Popper’s philosophy of science (Popper, 1963) — where a theory or model should only be considered “scientific” if there exists a means for that theory to be proven wrong — then we should accept capital models only insofar as there may be evidence that could be gathered that could cause us to reject the model results. Unfortunately in the case of operational risk, the extrapolation problem demonstrates how volatile and unreliable estimates of the 99.9th percentile of total annual losses can be. Due to the lack of data for

estimating high quantiles of the loss distributions, we can neither prove nor disprove that the capital estimates are valid to a reasonable level of accuracy. Nor can we prove or disprove that a method of extrapolating beyond the observed data is valid in the operational risk context, as we are not in a position to perform repeatable studies of loss distribution shapes in the range of our capital estimates.

In addition, the need for supplementary external data and possibly longer time series to better meet the challenge of gathering 1,000 years of banking data (or pooling data from 1,000 banks) introduces other challenges in the area of model validation. It expands the scope of the exercise beyond validation of the capital calculation process, to including as well the validation of the inputs (data quality) and the processing of the data prior to the estimation of the model. In particular, it must assess the use of any filtering and scaling methodologies for combining internal and external data in order to ensure that the external data is representative of the risks faced by the institution. Such tests only add to the data requirements of model validation. When including older data, particularly data older than the 5-year history requirement, the validation process should additionally include within its scope a verification of the relevance of this data given the changes that may have occurred in the institution's organization, such as mergers or divestments.

Considering the demands on the model validation process and the amount of data available, very little is afforded to the model validator except common sense. Of course, classical techniques used to validate output could still be used for lower percentiles of the loss distribution, where the model results are sometimes used for management purposes in day-to-day decision making, rather than for computing capital. However, the risk that the validation process turns into a sterile regulatory compliance exercise remains high, as does the risk of trying to fit unadapted techniques used in other model validation fields to the specific issues and challenges of operational risk.

In our view, there is room for one or more external operational risk validators, whose role it would be to benchmark capital models against each other with the goal of ensuring that best practices are encouraged, if and when they are applicable to the institution. The validation process employed by these external validators should be principles-based rather than rules-based, especially when applied across several institutions. This is necessary in order to preserve the differences in the way models are designed and implemented, and to not stifle innovation in the field. Most importantly, we feel that validation should place the model development process in the forefront of its methodology, focusing on issues of robustness. This aspect should validate that the proper tools and resources are employed to identify and develop innovative methodologies, and to continuously improve the model as the volume of data increases and the knowledge of how to solve some of the challenges and pitfalls exposed in this paper improves.

7 Summary

In this paper, we have demonstrated how the regulatory challenge of measuring operational risk at the 99.9 percentile and the consequential need to extrapolate introduces concerns regarding the reliability of results. The actual capital estimate should be "handled with care" due to instability of estimates and sensitivity of models to extreme data points. The difficulties that we point out naturally arise from dealing with heavy-tailed data (such as the dominance of sums and of mixtures). In this paper, we have deliberately used the lognormal as the "lightest" of the heavy-tailed distributions typically used in operational risk modeling, to illustrate most of our points. For heavier-tailed distributions such as the Pareto or log-logistic, the effects may be even more dramatic than what we have shown. In addition, the use of assumptions that cannot be verified (on mechanisms generating high severity events, on hard limits to incident severity, on differences in mitigation and control), should also raise questions as to the reliability of the result.

The challenges and pitfalls facing operational risk measurement at the higher percentiles thus thrusts the industry into a dilemma: on one hand, there is a need to hedge against low frequency / high severity events to protect the overall banking system, as demonstrated by the current banking crisis. On the other hand, according to the current regulations, we must place a great deal of confidence in the figures that are used to establish this hedge. The choice may also be posed by asking whether it is preferable to ensure partial security of the bank based on reliable data, or try to establish complete security by imposing a rigorous measurement criterion that must be based on unreliable statistics. If we pursue the goal of estimating 99.9th percentile of annual total losses, it may create only a perception of security, indeed very possibly a false one. Moreover, adopting models that have such a high degree of uncertainty to determine a regulatory capital adequacy standard will almost surely produce an uneven (and probably unfair because mostly related to random effects) distribution of capital among the various banks, creating an uneven playing field.

What could be therefore the way forward, to ease the severity of such a conclusion, without jeopardizing the progress that we have witnessed over the past several years in the area of operational risk modeling and management? Indeed, attempts at modeling operational risk have spread so-called "risk culture" to areas of the bank that had no familiarity with risk modeling techniques such as central functions. It has raised the attention on operational risk issues from line management to the board level and has developed instruments giving managers the ability to "read" levels of losses and understand causal analysis. None of these achievements should be put to waste, or efforts at modeling operational risk stopped.

Certainly, an immediate, albeit partial, way of overcoming the issue of data sufficiency is

to pursue and reinforce the effort of pooling data together from across institutions. Currently, data consortia such as ORX are beginning to fulfill this need, as its various working groups are building innovative models for data homogeneity and scaling analysis, capital benchmarking, and data quality and standards. The rigorous submission criteria of ORX have also given banks incentive to invest in additional features to their internal databases to establish proper mapping facilities for ORX criterion. In addition, the research conducted by its Analytic Agent on the data has told us much about its quality, and has contributed to the improvement of the reporting standards.

Another way to overcome data sufficiency is to better understand the causal links between process failures and operational losses, to help create a dynamic model of the loss generating mechanisms. Such models would increase the confidence we place on the overall validation process, in particular the validity of extrapolations beyond the data. In addition, such models would also improve the systematic use of scenarios to evaluate the frequency and severity of high impact events. Furthermore, such models could be of use to risk managers to better understand how to invest in improved processes and assets in order to reduce the risk, and hence the overall capital charge. We feel that this link to risk management and decision-making is sorely missing from the current measurement-based approaches.

In the medium term, we advocate lowering expectations about risk measurement by relaxing the supervisory standard. Lowering the soundness standard to, say, the 95 percentile would greatly increase the reliability of the model output. At these more modest levels, banks' efforts could be concentrated on improving the robustness of the model outputs and on their wider use in day-to-day managerial decisions. As the techniques to validate the reliability of the outputs grow more sophisticated, and the volume and quality of the data improves, models could be later reviewed and upgraded to include gradually higher confidence levels.

This however would not solve the regulatory concern for a 99.9 confidence interval. To reach this level, we suggest a more straightforward extrapolation method that is simple and unambiguous for banks to apply. For instance, scaling the capital calculated by multiplying the result of LDA at a lower confidence level by a fixed factor determined by the regulators is a possible solution. In this way, we try to keep capital numbers at a similar safety level but reduce the amount of uncertainty in the final output to less than 15% of the uncertainty of numbers estimated directly at the higher confidence level (see Appendix for details). This would provide more objective guidance for banks to assess their capital requirements, improve the robustness of the models, and reduce the level of uncertainty in the capital estimates. By doing so, it would also increase the perception of fairness across banks as the assessed capital numbers should be more uniform for similar institutions.

As a final suggestion, we also envision an improved model for the Standardized Approach

(TSA) for operational risk. Currently, TSA assesses operational risk capital by applying a fixed multiplier to the gross income of each business line in the bank, and summing the results. Improvements to this crude estimation procedure could be made based on scaling models of the type that have been developed in the industry (see, e.g., Cope and Labbi (2008)), that would take into account exposure indicators other than income that may be more relevant to the various business lines, as well as regional differences. An improved TSA would again reduce variability in capital estimates and produce similar capital estimates among similar banks, and would also tie the choice of scaling parameters to empirical studies of risk exposures based on cross-industry loss data.

References

- Arztner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999), ‘Coherent measures of risk’, *Mathematical Finance* **9**, 203–228.
- Asmussen, S. (2003), *Applied probability and queues*, Springer, Berlin.
- Böcker, K. and Klüppelberg, C. (2005), ‘Operational VaR: a closed form approximation’, *Risk Magazine* pp. 90–93.
- Braceras, R. M. (2004), ‘Late trading and market timing’, *The Review of Securities and Commodities Regulation* **37**(7), 61–70.
- Castillo, E., Hadi, A. S., Balakrishnan, N. and Sarabia, J. M. (2004), *Extreme value and related models with applications in engineering and sScience*, Wiley.
- Cope, E. and Antonini, G. (2008), ‘Observed correlations and dependencies among operational losses in the ORX consortium database’, *Journal of Operational Risk* **3**(4), 47–76.
- Cope, E. and Labbi, A. (2008), ‘Operational loss scaling by exposure indicators: evidence from the ORX database’, *Journal of Operational Risk* **3**(4), 25–46.
- de Haan, L. (1990), ‘Fighting the arch-enemy with mathematics’, *Statist. Neerlandica* **44**, 45–67.
- de Haan, L. and de Ronde, J. (1998), ‘Sea and wind: multivariate extremes at work’, *Extremes* **1**(1), 7–45.
- Degen, M., Embrechts, P. and Lambrigger, D. D. (2007), ‘The quantitative modeling of operational risk: between g-and-h and EVT’, *Astin Bulletin* **37**(2), 265–291.

- Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997), *Modeling extremal events for insurance and finance*, Springer.
- Embrechts, P. and Puccetti, G. (2008), ‘Aggregating operational risk across matrix structured loss data’, *Journal of Operational Risk* **3**(2), 29–44.
- Frachot, A., Moudaoulaud, O. and Roncalli, T. (2004), Loss distribution approach in practice, in M. Ong, ed., ‘The Basel Handbook: A Guide for Financial Practitioners’, Risk Books.
- Kiefer, J. (1967), ‘On Bahadur’s representation of sample quantiles’, *Ann. Math. Statist.* **38**, 1323–1342.
- Lindgren, G. and Rootzén, H. (1987), ‘Extreme values: theory and technical applications (with discussion)’, *Scandinavian Journal of Statistics* **14**, 241–279.
- Mignola, G. and Ugocioni, R. (2005), ‘Tests of extreme value theory’, *OpRisk & Compliance* pp. 32–35.
- Mignola, G. and Ugocioni, R. (2006), ‘Sources of uncertainty in modeling operational risk losses’, *Journal of Operational Risk* **1**(2), 33–50.
- Nešlehová, J., Embrechts, P. and Chavez-Demoulin, V. (2006), ‘Infinite mean models and the LDA for operational risk’, *Journal of Operational Risk* **1**(1), 3–25.
- ORX (2009), ‘Operational Riskdata eXchange database’.
URL: <http://www.orx.org>
- Panjer, H. (2006), *Operational risk: modeling analytics*, Wiley Series in Probability and Statistics.
- Popper, K. (1963), *Conjectures and refutations*, Routledge, New York.

Appendix: Uncertainty in Scaled Quantile Estimates: an Illustration

In this section, we illustrate the reduction in the level of uncertainty around the a scaled capital estimate, based on a fixed multiplicative factor times a lower quantile estimate such as the 95th percentile, from a direct quantile estimate of the 99.9th percentile. Suppose that the loss distribution is Pareto-distributed according to (3). From the asymptotic representation of the

standard error (2), we find that the standard error of a quantile estimate at level τ , $\hat{F}^{-1}(\tau)$, based on n independent samples, equals

$$\sigma(\hat{F}^{-1}(\tau)) = \frac{1}{\alpha}(1 - \tau)^{-(\alpha+1)/\alpha} \sqrt{\frac{\tau(1 - \tau)}{n}}.$$

We shall compare the standard error of a scaled lower quantile estimate with the standard error of a higher quantile estimate. Let τ_1 and τ_2 represent the two quantile levels, with $\tau_1 < \tau_2$, and let the scaling factor r equal the ratio of the true quantile levels $F^{-1}(\tau_2)/F^{-1}(\tau_1) = ((1 - \tau_1)/(1 - \tau_2))^{1/\alpha}$. We choose this value for r so that the expected value of the scaled quantile estimate equals the higher quantile level to which we want the capital value to correspond. In practice, of course, we would need to estimate the value of r , which would increase the error of the scaled quantile estimate, but we shall just assume here that the value of r is a given constant. Then the ratio

$$\frac{r \cdot \sigma(\hat{F}^{-1}(\tau_1))}{\sigma(\hat{F}^{-1}(\tau_2))} = \sqrt{\frac{\tau_1(1 - \tau_2)}{\tau_2(1 - \tau_1)}}, \quad (5)$$

which, interestingly, is always less than one, and does not depend on the value of α . When, for example, $\tau_1 = 0.95$ and $\tau_2 = 0.999$, then (5) equals 0.14, i.e., the error of a scaled estimate of the 95th percentile is less than 15% of the error of a direct estimate of the 99.9th percentile. Of course, we face the problem of choosing r appropriately, but if a good choice is made, we can make substantial reductions in the amount of uncertainty in the capital estimates without reducing capital levels than in the current situation.