BOJ Reports & Research Papers

December, 2007

Paper Series of Risk Management in Financial Institutions

The Effect of the Choice of the Loss Severity Distribution and the Parameter Estimation Method on Operational Risk Measurement*

– Analysis Using Sample Data –

Financial Systems and Bank Examination Department
Bank of Japan

* This is an English translation of the Japanese original released in June 2007.

Table of Contents

The Effect of the Choice of the Loss Severity Distribution and the Parameter Estimation Method on Operational Risk Measurement
– Analysis Using Sample Data –

Atsutoshi Mori*, Tomonori Kimata*, and Tsuyoshi Nagafuji*

(Abstract)

A number of financial institutions in Japan and overseas employ the loss distribution approach as an operational risk measurement technique. However, as yet, there is no standard practice. There are wide variations, especially in the specifications of the models used, the assumed loss severity distribution and the parameter estimation methods.

In this paper we introduce a series of processes for the measurement of operational risk: estimation of the loss severity distribution: estimation of the loss distribution and assessment of the results. For that purpose, we present an example of operational risk quantification for a sample data set that has the characteristics summarized below.

We use a sample data set extracted and processed from operational risk loss data for Japanese financial institutions. The sample data set is characterized as having 'stronger tail heaviness' than data drawn from a lognormal distribution, which is often used as a loss severity distribution.

By using this data set, we analyzed the effect on risk measurement of assumptions about the loss severity distributions and the effect of the parameter estimation methods used.

We could not find any distribution or parameter estimation method that is generally best suited. However, by analyzing the measurement results, we found that a more reasonable result could be obtained by: 1) estimating the loss severity distribution separately for low-severity and high-severity loss portions; and 2) selecting an appropriate parameter estimation method.

* Financial Systems and Bank Examination Department, Bank of Japan
E-mail: post.fsbe65ra@boj.or.jp

We appreciate helpful comments from Hidetoshi Nakagawa (Tokyo Institute of Technology).

1. Introduction

Many financial institutions in Japan and overseas are measuring their operational risk in order to manage it.[1] They use quantification to better understand their operational risk profiles and to estimate the economic capital for operational risk.[2]

Those financial institutions often face the following challenges in managing their operational risk through quantification:

1) There are challenges associated with the absence of any well-established practical technique for operational risk measurement.[3] For example, as different measurement techniques give significantly different quantification results, it is difficult to use them as objective standards for risk capital allocation and for day-to-day management. It is necessary to share an understanding of the characteristics of several major measurement techniques and of the differences in the risk amounts calculated.

2) There are challenges associated with the paucity of internal loss data. In this regard, Japanese financial institutions face two challenges. First, few institutions have collected enough internal operational loss data. Second, it is very difficult for institutions to find an external operational risk database suitable for them.

In this paper, we aim to develop a process for operational risk measurement that contributes to financial institutions' efforts to measure their operational risk and enhances their operational risk management.

To that end, we perform a comparative analysis of the characteristics, advantages, and disadvantages of various techniques used in many financial institutions in terms of their applicability to actual loss data and in terms of the validity of the measured amounts of risk. We measure operational risk based on operational risk loss data collected from financial institutions in Japan by using various risk measurement techniques.[4]

To understand this paper, readers should be aware of several issues relating to the sample data analyzed and the measurement techniques described. First, the sample data used in this paper are restricted in the sense that data on higher severity losses with low frequency (low-frequency high-severity losses) may not have been collected.[5] This is inevitable when measuring operational risk. In addition, although in this paper we mainly use proven measurement techniques that have already been widely used by financial institutions (including the loss distribution approach),[6] it is quite likely that other superior techniques are available. In addition, because operational risk

[1] The term operational risk, as used herein, is defined as the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events, including legal risk (the risk that includes exposure to fines, penalties, or punitive damages resulting from supervisory actions, as well as private settlements) but excluding strategic risk (the risk of any loss suffered as a result of developing or implementing improper management strategy) and reputational risk (the risk to financial institutions of losses suffered as a result of a deterioration in creditworthiness due to the spread of rumors).

[2] See the Study Group for the Advancement of Operational Risk Management [2006].

[3] See the Study Group for the Advancement of Operational Risk Management [2006].

[4] The data record information about each loss resulting from an operational risk event is the amount of each loss and the date when it occurred.

[5] The characteristics of the data used herein are described in Section 4.

[6] A summary of loss distribution approach is provided in Section 3.

measurement techniques remain under development, other new and better techniques may be developed in future. Moreover, the techniques preferred in this paper may not necessarily be appropriate for data that exhibit different operational risk characteristics.

The paper is organized as follows: The next section surveys examples of earlier studies of the measurement of operational risk. In Section 3 we summarize the risk measurement framework. In Section 4 we outline the characteristics of the sample data used in this paper. In Section 5 we estimate the loss severity distribution by using various methods and compare the results obtained from those methods. In Section 6, we review the aforementioned processes and summarize the practical insights gained about operational risk measurement and discuss outstanding issues.

Matters relevant to the subject that may help to illuminate the discussion are provided as supplementary discussions. In Appendix 1, we explain the relationship between confidence intervals in risk measurement and the range of the loss data that may affect the amount of risk. In Appendix 2, we provide an explanation of technical terms and issues.

2. Examples of Earlier Studies

There are a number of analyses of operational risk measurement that use loss data. To our knowledge, the only publicly available analysis performed in Japan is the one by the Mitsubishi Trust & Banking Corporation's Operational Risk Study Group [2002]. There are a number of overseas studies, including those by de Fontnouvelle *et al.* [2003], de Fontnouvelle *et al.* [2004], Chapelle *et al.* [2004], Moscadelli [2004], and Dutta and Perry [2006]. Below, we summarize these papers from the viewpoint of the data and the measurement techniques used.

2.1. Data Used for Measurement

With the exception of the study by de Fontnouvelle *et al.* [2003], who used commercially available loss data, all the studies used internal loss data from a single or several financial institutions (data on actual losses collected from the financial institution(s)).

Leaving aside the study by de Fontnouvelle *et al.* [2003], the Mitsubishi Trust & Banking Corporation's Operational Risk Study Group [2002] and Chapelle *et al.* [2004] used data from a single financial institution, whereas Moscadelli [2004], de Fontnouvelle *et al.* [2004], and Dutta and Perry [2006] used loss data from a number of financial institutions (ranging from six to 89 banks).

Of the authors that used internal loss data from more than one financial institution, de Fontnouvelle *et al.* [2004] and Dutta and Perry [2006] measured risk on an individual institution basis. Moscadelli [2004] measured risk after having consolidated data from all financial institutions.

For all studies, loss data were classified into several units of measurement based on the event type, business line, or both.[7] Then, operational risks by measurement unit were quantified.

2.2. Techniques Used for Measurement

In all studies that used the loss distribution approach to measure risk, it was found that measurement results depend significantly on the shape of severity distribution assumed. In all studies, extreme value theory (the Peak Over Threshold (POT) approach) was used to develop the quantification model,[8] taking into account the tail heaviness of the operational risk loss distribution. de Fontnouvelle *et al.* [2004], Chapelle *et al.* [2004], and Moscadelli [2004] favored the use of extreme value theory (the POT approach). However, Dutta and Perry [2006] criticized this method on two grounds. First, the method yields an unreasonable capital estimate. Second, the measured amounts of risk depend heavily on the thresholds used. Thus, Dutta and Perry [2006] advocated the use of a distribution with four parameters, which allows for a greater degree of freedom.[9]

With regard to the parameter estimation method used for the severity distribution, the Mitsubishi Trust & Banking Corporation's Operational Risk Study Group [2002] demonstrated that the amount of risk depends significantly on the estimation method applied. However, in other studies, only one technique (typically maximum likelihood estimation) was used. Moreover, there was no comparison or evaluation of the calculated amounts of risk obtained on the basis of different parameter estimation methods.

In this paper, first, we quantify risk by applying a single severity distribution to the full sample data set. Second, we measure risk by applying a compound distribution to the sample data set. As we explain later, use of the compound distribution involves estimating two different distributions, one above and one below a certain threshold, after which the distributions are consolidated. In using the compound distribution, we applied the concept of extreme value theory (the POT approach), as used in existing studies, to low-frequency, high-severity loss data.

3. Summary of the Loss Distribution Approach

In this section, we describe some basic techniques and concepts used in this paper. First, we introduce the framework for the loss distribution approach, which is used in this paper. Second, we explain the loss distribution approach (parametric method) used for analysis, and then we explain the nonparametric method, which has been adopted as a

---

[7] They all used the Basel II business lines (e.g., corporate finance, retail banking) and event types (e.g. internal fraud, clients, products & business practices).

[8] Extreme value theory is a theory that addresses distributions formed by extremely large values (the extreme value distribution). The POT approach is a method used to estimate the extreme value distribution based on the proposition that "if the threshold is set at a sufficiently high level, the distribution of amounts in excess of the threshold can be approximated by a generalized Pareto distribution" (the Pickands–Balkema–de Haan theorem). When the POT approach is used for the measurement of operational risk, the threshold for the loss data is set at an appropriate level, and it is assumed that the amount of data in excess of the threshold (i.e., the tail) forms a generalized Pareto distribution. See Morimoto [2000] for further details.

[9] Introduced by Hoaglin *et al*. [1985] as a g–and-h distribution.

benchmark for evaluating the risk measurements based on the parametric method. Third, we discuss our justification and the conditions required for using the nonparametric method as a benchmark in this paper.

3.1. Framework for Loss Distribution Approach

In this paper, we define the amount of operational risk as value at risk (VaR),[10] that is, the amount of risk based on a confidence interval of 100α%, as the 100α percentile point of the loss distribution, i.e., the distribution of the total amount of all the loss events that occur during the risk measurement period.

The estimated loss distribution combines the loss frequency distribution (the probability distribution of the number of times a loss occurs during the risk measurement period) and the loss severity distribution (the probability distribution of the amount of loss incurred per occurrence).

We assume a risk measurement period of one year and confidence intervals of 99% and 99.9%. We use Monte Carlo simulation (hereafter, simulation) to estimate the loss distribution.[11]

The amount of risk is estimated by using the following process.

1) Estimation of the Loss Frequency Distribution

The distribution of *N*, the number of losses during the risk measurement period of one year (the loss frequency distribution), is estimated. We assume that *N* follows a Poisson distribution, for which we assume parameters based on the annual average number of loss events.[12]

2) Estimation of the Loss Severity Distribution

Having estimated the distribution, we estimate $X_i, (i = 1,2,..., N)$, which represents the amount of loss per occurrence of the loss event (the severity distribution). Broadly, the methods used to estimate severity distributions can be classified into two types: one is parametric methods, in which a particular severity distribution (e.g., lognormal or Weibull) is assumed, and the other is nonparametric methods, in which no particular distribution is assumed. We assume that the severity of each loss, represented by X_i, is independent and identically distributed (i.i.d.). We also assume that the number of loss events and the severity of each loss, represented by N and X_i respectively, are independent of each other.

[10] We use the VaR, which is most largely used in practice for operational risk quantification.

[11] Other methods of calculating the total loss distribution without using a simulation include a method based on Panjer's recurrence equation and the fast Fourier transformation, which are well known. See Klugman *et al*. [2004] for details.

[12] The probability function of the Poisson distribution is $f(x) = \frac{e^{-\lambda}\lambda^x}{x!}$. The expected value is λ, which is estimated by equating this with the annual average number of loss events.

3) Calculation of the Loss Amount

Using the loss frequency distribution estimated in 1) above, the number of annual losses (N) is derived. Then, the severity of losses for N occurrences, represented by $(X_1, X_2, \cdots, X_N)$, is derived from the severity distribution estimated in 2) above. Then, the total amount of loss for the risk measurement period of one year, represented by S , is calculated as follows:

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

4) Calculation of the VaR by Simulation (from a Trial of K Times)

Step 3) is repeated K times to calculate the severity for K trials, i.e., $S_{(1)}, S_{(2)}, ..., S_{(K)}$, which are arranged in ascending order as S_i $(S_1 \le S_2 \le \cdots \le S_K)$. The amount of risk is defined as:

$$\begin{aligned} VaR(\alpha) &= S_i, \quad \frac{i-1}{K} \le \alpha < \frac{i}{K}, \quad i = 1,2,...,K \\ &= S_{[\alpha K+1]}, \end{aligned}$$

where [x] represents the largest integer that is smaller than x.

For example, if K = 100,000 and α = 0.99, the amount of risk is $VaR(0.99) = S_{99001}$; i.e., the 1000th severest in terms of the total amount of loss.

3.2. Loss Severity Distribution Estimation Methods (Parametric and Nonparametric)

3.2.1. Parametric Methods

The parametric method assumes a particular severity distribution. In this paper, the distributions assumed are the lognormal, the Weibull, and a generalized Pareto distribution.[13] The estimation methods used are the Method of Moments (MM) (with a probability-weighted method of moments (PWM) being used for the generalized Pareto distribution), Maximum Likelihood Estimation (MLE), and Ordinary Least Squares (OLS).[14]

3.2.2. Nonparametric Methods

Unlike the parametric method, the nonparametric method derives a loss amount at random from loss data to perform a simulation without assuming any particular severity distribution.

---

[13] In general, distributions that can capture tail heaviness are chosen for severity distributions in operational risk quantification, as tail heaviness is a characteristic of operational risk. There are other types of distribution that can be used for loss amounts, such as the gamma distribution and the generalized extreme value distribution.

[14] See (1) and (2) in [Appendix 2] for the characteristics and shapes of distributions used for loss severity and the concepts and characteristics of the parameter estimation methods in this paper.

Given L items of loss data, in this paper, we arrange the data items in ascending order of loss amounts as $X_i \ (X_1 \leq X_2 \leq \cdots \leq X_L)$. We then define the following function for X that yields a particular loss (in which p represents a probability; $0 < p < 1$):

$$
\begin{aligned}
X(p) &= X_i, \quad \frac{i-1}{L} \leq p < \frac{i}{L}, \quad i = 1,2,\ldots,L \\
&= X_{[Lp+1]}
\end{aligned}
$$

3.3. The Nonparametric Method as a Benchmark

We treat the estimated risk based on the nonparametric method, which assumes no particular loss severity distribution, as a benchmark for the risk estimated from using the parametric method.

Because of the small number of data points in the sample data set used, this benchmark does not necessarily represent a conservative amount of risk.[15]

4. Data

We use the observations corresponding to the 774 largest loss amounts, obtained from operational risk loss data on Japanese financial institutions over a 10-year period from January 1994 to December 2003. Hence, from the viewpoint of a single financial institution, the sample data used can be considered as a loss database that comprises loss data for other banks (external loss data) in addition to its own internal loss data.

To determine the characteristics of the sample data set used for operational risk measurement, we evaluate the 'tail heaviness' of the sample distribution.[16] As shown in [Table 1], the distribution of the sample data exhibits heavier tails than those of the lognormal distribution.

To evaluate tail heaviness, we compare the percentiles of the two distributions for the same loss amount: the distribution of the sample data and the lognormal distribution estimated from the sample data; the latter is often used for severity distributions. The comparisons are based on various loss amounts.

We adopt the following process to evaluate tail heaviness.

1) The sample data were arranged in ascending order of loss amount as $X_i \ (X_1 \leq X_2 \leq \cdots \leq X_N)$ to calculate the average logarithm value (μ) and the standard deviation (σ), which were used to normalize the data as follows:

$$
Y_i = \frac{\log X_i - \mu}{\sigma}
$$

[15] These issues are discussed in Appendix1.

[16] In this paper, for two distribution functions $F(x)$ and $G(x)$, if there exist some amount represented by x_0 such that for any $x > x_0$, $1 - F(x) > 1 - G(x)$, i.e., $F(x) < G(x)$, then we define the distribution represented by $F(x)$ has a heavier tail than does the distribution represented by $G(x)$.

2) For the distribution function for the normalized sample values Y_i , each Y_i is defined as follows:

$$S_N(Y_i) = \frac{i-0.5}{N},\ i = 1,2,\cdots,N$$

3) The standard normal distribution function is denoted by $F(x)$.

4) The values of the distribution functions defined in 2) and 3) above are compared to identify the tail heaviness of the sample data. In this context, we set $x_n = 0.5n$, $(n = 1,2,\cdots,8)$ and calculated $S_N(Y^n)$ by using $F(x_n)$ for each point, where Y^n is the smallest value of Y_i that satisfies $x_n \le Y_i$ (i.e., Y_i is the minimum value that is at least as large as x_n). In this context, we assume that $S_N(x_n) \le S_N(Y^n)$ represents an appropriate definition of $S_N(x_n)$, because the distribution function is monotonically nondecreasing.

These calculations are summarized in [Table 1]. The sample data have heavier tails than those of the lognormal distribution, that is, for all loss amounts, if $x_n \ge 1.5$, the following relationship holds:

$$(S_N(x_n) \le)\ \ S_N(Y^n) < F(x_n)$$

[Table 1] Comparative Verification of the Tail Heaviness of the Severity Distribution

n	x_n	$S_N(Y^n)$ (A)	$F(x_n)$ (B)	Difference (B-A)
1	0.5	0.76421	0.69146	-0.07275
2	1	0.84690	0.84134	-0.00555
3	1.5	0.90504	0.93319	0.02815
4	2	0.94767	0.97725	0.02958
5	2.5	0.97222	0.99379	0.02157
6	3	0.98385	0.99865	0.01480
7	3.5	0.99160	0.99977	0.00817
8	4	0.99677	0.99997	0.00320

In what follows, we evaluate the validity of each method applied to the sample data, which have the tail heaviness shown in this section, both on the basis of goodness of fit for the tails of the distribution and on the basis of the amount of risk.

5. Measurement Results and Analysis

In this section, we evaluate the results of measuring operational risk based on the sample data set discussed in Section 4 by using the techniques described in Section 3.

First, in subsection 5.1, we calculate and analyze the amount of risk by assuming a single severity distribution. Then, in subsection 5.2, we calculate and analyze the amount of risk by assuming a compound severity distribution. This is done to improve the goodness of fit in part of the distribution, which tends to be poor when using a single distribution.

In both cases, we use the quantification result of the nonparametric method as a benchmark. In addition, we use a PP plot or a QQ plot, as applicable, to assess the goodness of fit of the assumed distribution of the loss data.[17, 18]

5.1. Methods that Assume a Single Severity Distribution

5.1.1. Quantification Method

To quantify the amount of risk, we applied three different single severity distributions to the whole data set and used three parameter estimation methods. The estimated parameters are shown in [Table 2].

The distributions used were the lognormal, the Weibull, and a generalized Pareto distribution. For parameter estimation, we used MLE, OLS, and MM (with PWM being used for the generalized Pareto distribution).[19]

---

[17] See [Appendix 2] 3 for an explanation of PP and QQ plots.

[18] We rely on a visual technique, such as inspecting the PP or the QQ plot to assess the fit of the tail, which has a great impact on the amount of risk. The widely known statistical techniques (such as the Kolmogorov–Smirnov test or the Anderson–Darling test) cannot fully assess the fitness in the tail of a very heavy-tailed dataset. See [Appendix 2] 4 for details.

[19] See [Appendix 2] 1 and 2 for the characteristics and shapes of distributions used for loss severity and the concepts and characteristics of the parameter estimation methods in this paper.

[Table 2] Results of Parameter Estimation When a Single Severity Distribution is Assumed

	Lognormal Distribution					
	MM		MLE		OLS	
Results of Parameter Estimation	μ	σ	μ	σ	μ	σ
	2.17	2.47	1.48	1.61	1.48	1.47
	Weibull Distribution					
	MM		MLE		OLS	
Results of Parameter Estimation	θ	p	θ	p	θ	p
	0.70	0.18	10.85	0.43	10.90	0.63
	Generalized Pareto Distribution					
	PWM		MLE		OLS	
Results of Parameter Estimation	β	ξ	β	ξ	β	ξ
	4.57	0.98	3.44	1.10	3.84	0.60

5.1.2. Results of Risk Measurement

The amounts of risk at confidence intervals of 99% and 99.9% calculated using the estimated parametric severity distributions and the nonparametric severity distribution are shown in [Table 3].

[Table 3] shows that the estimated amount of risk depends greatly on the distribution assumed and the parameter estimation method chosen. When the lognormal or Weibull distribution is assumed, the estimated amount of risk is high for MM. In contrast, under MLE and OLS, the amount of risk is small. The ratio between the amounts of risks estimated under MM and under OLS at 99% confidence interval is 42:1 ([Table 3] (A)) for the lognormal and 65:1 ([Table 3] (B)) for the Weibull. At 99.9% confidence interval, the corresponding ratios are 99:1 ([Table 3] (C)) and 190:1 ([Table 3] (D)). When the generalized Pareto distribution is assumed, the amount of risk obtained under MLE is high and the amount of risk under OLS is low. At confidence intervals of 99% and 99.9%, the ratios between the estimates under MLE and OLS are 30:1 ([Table 3] (E)) and 125:1 ([Table 3] (F)), respectively.

[Table 3] Amount of Risk When a Single Severity Distribution is Assumed

Confidence Interval		99% (α)	99.9% (β)	$(\beta)/(\alpha)$
Lognormal Distribution	MM	74.9 <1>(0.75)	(C) 99:1 272.6 <3>(1.4)	3.6
	MLE	2.5 (0.025)	4.4 (0.023)	1.8
	OLS	1.8 <9>(0.018)	2.8 <9>(0.015)	1.6
Weibull Distribution	MM	105.8 <2>.(1.1)	(D) 190 :1 351.7 <4>(1.9)	3.3
	MLE	3.9 (0.039)	5.0 (0.026)	1.3
	OLS	1.6 <9>(0.016)	1.8 <9>(0.010)	1.1
Generalized Pareto Distribution	PWM	26.8 <5>(0.268)	(F) 125 :1 255.4 <7>(1.348)	9.5
	MLE	(E) 30 :1 54.1 [illegible]>(0.541)	686.1 <8>(3.622)	12.7
	OLS	1.8 <9>(0.018)	5.5 <9>(0.029)	3.0
Nonparametric Method		100.0 (1.00)	189.4 (1.00)	1.9

Notes:

1) The amount of risk is the relative value indexed to the value based on the nonparametric method (at 99% confidence), which represents 100.

2) The figures in brackets represent the scaling factors for the amounts of risk against the benchmark at each confidence interval.

3) The number of trials is 100,000.

5.1.3. Assessment and Discussion of Quantification Results

Next, we benchmarked the quantification results based on the parametric method against the results based on the nonparametric method. When the lognormal or Weibull distribution is assumed, if MM is used for parameter estimation, the amount of risk is at least as high as the benchmarks. At the 99% confidence interval, the ratios of the parametrically estimated amount of risk to the benchmark (nonparametorically estimated amount) are 0.75:1 ([Table 3] <1>) and 1.1:1 ([Table 3] <2>) for the lognormal and Weibull distributions, respectively. At the 99.9% confidence interval, the corresponding figures are 1.4:1 ([Table 3] <3>) and 1.9:1 ([Table 3] <4>).

When MLE or OLS is used, in all cases, the parametrically estimated amount of risk is less than 5% of the benchmark, thus falling well below it.

When the generalized Pareto distribution is assumed, at the 99% confidence interval, for PWM and MLE, the ratios are 0.27:1 ([Table 3] <5>) and 0.54:1 ([Table 3] <6>), respectively, and thus fall below the benchmark. At the 99.9% confidence interval, the corresponding ratios are 1.35:1 ([Table 3] <7>) and 3.62:1 ([Table 3] <8>), and thus exceed the benchmark. By contrast, when OLS is used, at both confidence invervals, the amount of risk is less than 3% of the benchmark ([Table 3] <9>).

The differences in the estimated risk amounts arising because of the distribution assumed or the parameter estimation method adopted are interpreted below.

1) Distribution Assumed

The variations between the results based on different distributions are caused by differences in the tail heaviness of the distributions. Among the severity distributions we used, it is generally known that the Weibull distribution is the least tail-heavy distribution, followed by the lognormal distribution, and then by the generalized Pareto distribution.[20]

2) Parameter Estimation Method

In our analysis, there are quite significant variations in the results because of differences in the parameter estimation method used. This means that, in our analysis, there is a substantial difference between the assumed distribution and the data. Unless there is a large deviation, a parametric distribution yields a similar approximation irrespective of the parameter estimation method used.

The PP and QQ plots confirm this. Using the PP plot for the lognormal severity distribution as an example, when MLE and OLS are used, although there is a reasonable goodness of fit in the central part (the body) of the distribution, on the right side of the distribution (in the tail), the loss amount declines, which leads to a difference between the estimates and the data. By contrast, if MM is used, although there is a large deviation in the data in the body, the deviation in the tail is smaller. In addition, according to the QQ plot, the deviation from the data, particularly in the tail, is larger under MLE and OLS than under MM (see [Table 4]).

[20] It is generally known that in terms of the degree of tail heaviness they are ranked in the following order: the generalized Pareto distribution, the lognormal distribution, the Weibull distribution (if $p<1$), the gamma distribution, and the Weibull distribution (if $p>1$). Of these, the generalized Pareto distribution has the heaviest tail; i.e., between each distribution function, $F_{GPD}(x), F_{LN}(x), F_{WB,p<1}(x), F_{GAM}(x), F_{WB,p>1}(x)$ and for x of a sufficiently large value, the equality

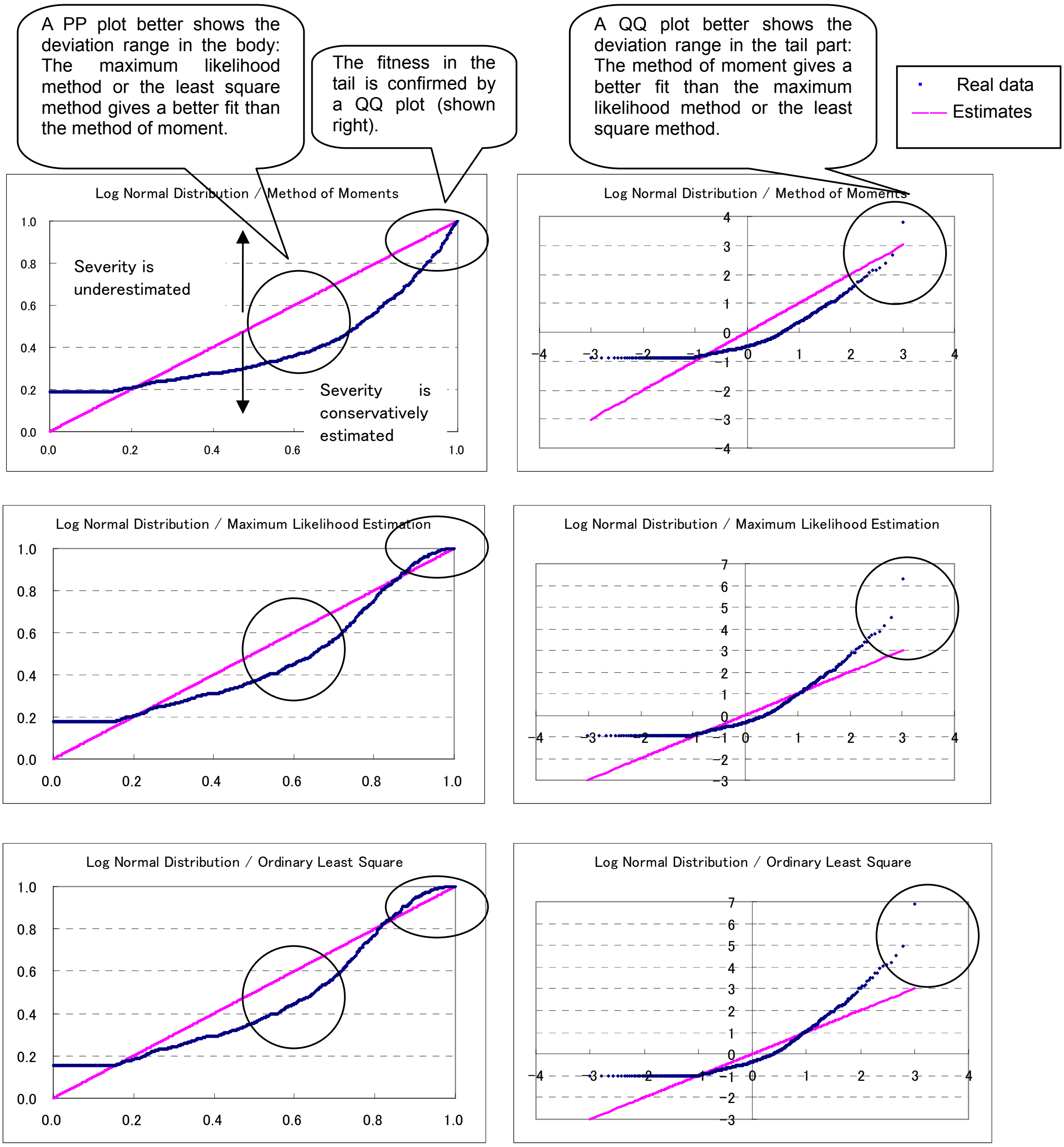
$$F_{GPD}(x) < F_{LN}(x) < F_{WB,p<1}(x) < F_{GAM}(x) < F_{WB,p>1}(x)$$

is true. In all cases, the shape parameter p of the Weibull distribution used to measure risk in this paper (see [Appendix 2] (1) for the parameter of the Weibull distribution) is less than unity.

[Figure 4] Fitness Assessment Using PP / QQ Plot Assuming a Single Distribution

The PP / QQ plots (*) are shown for three types of parameter estimation method assuming a log-normal severity distribution.

＜PP Plot＞ ＜QQ Plot＞



*In the QQ Plot both X - and Y axes standardize and represent the average and standard deviation as 0 and 1, respectively, for the estimates and the log values of the data based on the assumed parameters (as with any QQ plot hereinafter).

As explained above, when it is difficult to fit a single parametric distribution to the whole data, the amount of quantified risk depends greatly on the parameter estimation method used, which determines which part of the data the estimated distribution fits well.

More precisely, when MM is used, the quantified risk amounts are at least as high as the benchmarks, which are the estimated risk amounts based on the nonparametric method. In this case, these estimates fit well in the tail, whereas there are large deviations between the two distributions in the body.

In contrast, when MLE and OLS are used, the quantified risk amounts are below the benchmarks. The two distributions fit well in the body, but underestimate the severity of loss in the tail.

The calculations above suggest that the goodness of fit in the tail of the distribution has a particularly marked effect on the estimated amount of risk. Therefore, to estimate the amount of risk, it is important to check goodness of fit to the data in the tail of the distribution before assuming the distribution. Then, one can choose a parameter estimation technique.

A parameter estimation technique that fits the tail well yields a better estimate of the amount of risk. This is because the amount of risk calculated is greatly affected by the tail. When we used a lognormal distribution for the sample data, MM seems to be a more appropriate method for parameter estimation than MLE or OLS.[21]

When there is a large deviation between the assumed distribution and the distribution of the data, even if the amount of risk calculated at a certain confidence interval is the same as at the benchmark, the amount of risk calculated at another confidence interval may not necessarily match the benchmark. For example, the generalized Pareto distribution, when estimated by using the PWM method, yields a measure of risk that is well below the benchmark at the 99% confidence interval, but yields a risk amount that is well above the benchmark at the 99.9% confidence interval.

It is difficult to find a single severity distribution that fits well throughout the range of the sample data from the body to the tail. That the estimated amount of risk depends greatly on the parameter estimation technique used, whatever distribution is assumed, confirms this.

For this reason, to improve goodness of fit in the tail, in the next subsection, we conduct an analysis based on the compound distribution.

[21] However, such a relationship between the parameter estimation techniques and the risk quantification results is not always stable, and depends on the distribution conditions of the data. For example, according to the Mitsubishi Trust & Banking Corporation's Operational Risk Study Group [2002], when the amount of risk calculated by using MLE exceeds the amount of risk calculated by using MM, the magnitude of the relationship between the parameter estimation techniques and the quantification results is reversed.

5.2. Methods that Assume a Compound Severity Distribution

5.2.1. Risk Measurement Method

To avoid the problems in fitting a single parametric distribution to the whole data, we use a compound severity distribution. This involves dividing the severity distribution into the body and the tail and assuming a different distribution for each part. These distributions are then combined into a single severity distribution (termed a compound distribution). A threshold is set for the loss amount,[22] and different distributions (one for the body and one for the tail) are estimated for values above and below this threshold value. These distributions are then consolidated into a single severity distribution (the compound distribution) and a Monte Carlo simulation is performed.

Details of this process are given below:

1) Setting the Thresholds

The minimum loss amount that exceeds the percentile point p when the loss data are arranged in ascending order is used as the threshold $T(p)$. Losses below and above the threshold are referred to as low-severity and high-severity losses, respectively.[23] The loss data are denoted by L_i, $(i=1,2,...,L)$, and the threshold is defined as follows:

$$
\begin{aligned}
T(p) &= L_i, \quad \frac{i-1}{L} \le p < \frac{i}{L}, \quad i=1,2,...,L \\
&= L_{[pL+1]}, \qquad [\mathrm{x}] \text{ represents the largest integral number not exceeding x.}
\end{aligned}
$$

2) Estimation of the Loss Frequency Distribution

As for the case of a single severity distribution, the loss frequency distribution is estimated. We assume a common loss frequency distribution for the body and the tail. This means that the total number of high-frequency low-severity and low-frequency high-severity losses during the year is represented by N, which is assumed to follow a Poisson distribution.

3) Estimation of the Severity Distribution

To estimate the distribution for the amount of loss per occurrence of a loss event $X_i, (i=1,2,...,N)$ (the loss severity distribution), the following process is adopted:

(i) Estimation of the Severity Distribution in the Body

The severity distribution for the body (for which the distribution function is $F_b(x)$) is estimated by using the full data set.[24]

---

[22] The threshold may be set at a certain amount of money as well as at a certain percentile point. In this paper, we use the latter approach.

[23] Three threshold levels are assumed: 90% ($p=0.9$), 95% ($p=0.95$), and 99% ($p=0.99$).

[24] We chose the lognormal distribution for the severity distribution in the body and MM for the parameter estimation technique. We numerically verified that, regardless of the point at which the threshold was set, neither the assumption about the distribution nor the chosen parameter estimation technique had any significant impact on the risk quantification results.

(ii) Estimation of the Severity Distribution in the Tail

The severity distribution in the tail (for which the distribution function is $F_t(x)$) is estimated by using the observed loss amounts that exceed the threshold. Three distributions, the lognormal, the Weibull, and a generalized Pareto distribution, are used for the tail, as was the case when a single distribution was used (see previous subsection). For parameter estimation for the tail, MLE, OLS, and MM (PWM for the generalized Pareto distribution) are used.

(iii) Compounding Distributions

The distributions estimated in (i) and (ii) are combined at the threshold to produce a single compound distribution (referred to as $F(x)$) after making adjustments to eliminate overlapping and gaps.

The percentile point of the threshold $T(p)$ of the distribution function for the body is represented by α, and the distribution functions for $F_b(T(p)) = \alpha$ and $F(x)$ are defined as follows:

$$F(x) = \begin{cases} \frac{p}{\alpha} F_b(x), & 0 \leq x < T(p) \\ p, & x = T(p) \\ p + (1-p)F_t(x - T(p)), & T(p) < x \end{cases}$$

This means that, for the distribution in the body, the value of the distribution function is scaled so that the area of the density function in the part below the threshold is equal to $100p\%$. For the distribution in the tail, the value of the distribution function is scaled so that the area of the density function in the part above the threshold is equal to $100(1-p)\%$.

This method embraces the concept of the extreme value method (POT approach),[25] in that two different distributions are combined to form a single distribution, but it does not strictly apply this method. This is because the generalized Pareto distribution did not fit well in the high-severity loss portion above any threshold value that we tried at 90%, 95%, or 99% confidence. For this reason, we did not apply the Pickands–Balkema–de Haan Theorem, which states that "the distribution of the observations in excess of a certain high threshold can be approximated by a generalized Pareto distribution."[26]

Instead, we first considered different threshold values without insisting on a statistical justification, and second, used a distribution for the tail other than the generalized

[25] See footnote 8.

[26] To apply extreme value theory (POT approach) strictly, it is necessary to verify whether the data above the threshold, i.e. the data that rank in the top $100(1-p)$ % if the threshold is set at the $100p$ % point from the bottom of the data, take a generalized Pareto distribution. Then, if such a condition is satisfied, first, a generalized Pareto distribution is assumed for those data that rank in the top $100(1-p)$ %, and second, another distribution is used for the remaining data (below the $100p$ % from the bottom). The parameter of each distribution is then estimated, and these are combined to form a single severity distribution, based on which the risk is measured.

Pareto distribution.

It is worth assuming a different distribution for different loss amounts to estimate the parameters for each distribution as a practical experiment. This is because there are different causes of high-frequency low-severity losses and low-frequency high-severity losses.[27]

The parameter estimation results for the tail obtained from a compound distribution are shown in [Table 5].[28]

[Table 5] Results of Parameter Estimation in the Tail When a Compound Severity Distribution is Assumed

Tail	Lognormal Distribution					
	MM		MLE		OLS	
Threshold	μ	σ	μ	σ	μ	σ
90%	5.51	1.98	4.41	1.96	4.41	2.02
95%	6.49	1.81	5.21	2.19	5.21	2.34
99%	8.66	1.38	7.72	1.77	7.72	2.19
Tail	Weibull Distribution					
	MM		MLE		OLS	
Threshold	θ	p	θ	p	θ	p
90%	106.27	0.267	231.07	0.426	218.11	0.571
95%	418.59	0.309	560.57	0.419	543.01	0.501
99%	6864.16	0.478	6018.15	0.477	7020.14	0.420
Tail	Generalized Pareto Distribution					
	PWM		MLE		OLS	
Threshold	β	ξ	β	ξ	β	ξ
90%	130.80	0.925	44.16	2.40	62.91	1.36
95%	380.68	0.888	74.69	2.93	151.09	1.54
99%	5334.73	0.643	567.35	2.99	1355.95	1.30

5.2.2. Results of Risk Measurement

The risk amounts quantified at confidence intervals of 99% and 99.9% when a compound distribution is used are shown in [Table 6]. As for the case of a single distribution, the estimated amount of risk based on the nonparametric method is used as a benchmark and is shown in the table. We do not report the results obtained from a generalized Pareto distribution under MLE because the estimates were implausibly large.

[27] See the Study Group for the Advancement of Operational Risk Management [2006].

[28] In all cases, we used the values calculated based on using the lognormal distribution and MM ($\mu = 2.17, \sigma = 2.47$) for the parameters in the body of the distribution.

For this distribution, at the 99% and 99.9% confidence interval, the maximum likelihood estimates were between 1,000 and 10,000 times larger than those obtained when using PWM. This is because the estimate of the shape parameter for the generalized Pareto distribution exceeded unity (which implies an extremely heavy tail).[29]

[29] See [Appendix 2] 1. for the parameters and characteristics of a generalized Pareto distribution.

[Table 6] Amount of Risk Assuming a Compound Loss Amount Distribution

(Conditions)

・Data: the high-severity loss portion of the sample data (at or in excess of the 90%, 95%, and 99% points) is assumed to be the tail.

・The tail (distribution): The lognormal distribution, the Weibull distribution, and the generalized Pareto distribution.

(The parameter estimation technique): MM (PWM for the generalized Pareto distribution), MLE, OLS.

・The body (distribution): the lognormal distribution.

(The parameter estimation technique): MM

・Number of simulations: 100,000

The difference in the volume of risk is small

Distribution in the tail	Parameter estimation technique	Estimated at a confidence interval of 99%				Estimated at a confidence interval of 99.9%			
		Threshold			Single distribution	Threshold			Single distribution
		90%	95%	99%		90%	95%	99%	
Lognormal Distribution	MM	93.1	100.9	120.4	74.9	297.7	316.1	314.0	272.6
	MLE	29.7	76.9	107.9	2.5	92.0	300.2	395.0	4.4
	OLS	35.5	115.9	262.2	1.8	114.8	527.0	1,412.2	2.8
Weibull Distribution	MM	120.7	126.2	140.1	105.8	346.2	328.7	318.9	351.7
	MLE	22.8	41.3	124.1	3.9	40.2	80.7	284.9	5.0
	OLS	9.1	22.2	213.1	1.6	13.1	37.7	563.1	1.8
Generalized Pareto Distribution	PWM	61.1	77.1	120.0	26.8	453.4	556.0	505.6	255.4
	MLE	n.a.	n.a.	n.a.	54.1	n.a.	n.a.	n.a.	686.1
	OLS	345.4	782.5	257.6	1.8	7,513.0	22,953.3	4,396.5	5.5
Nonparametric Method					100.0				189.4

・Amount of Risk (single distribution; the relative value indexed to the value based on the nonparametric method (at 99% confidence), which represents 100)

Boundary point	90%	95%	99%
Value of the boundary point	0.04	0.10	0.85
Number of data pieces in the body	696	735	766
Number of data pieces in the tail	78	39	8

At the confidence interval of 99%: the same risk amount as MM.
At the confidence interval of 99.9% MM produced a risk amount equal to approx. 1.5 times the risk amount produced by the nonparametric method.

Compared with the single distribution analyzed in the previous subsection, a compound distribution yielded smaller variations in the amount of risk depending on the distribution assumed and on the parameter estimation technique chosen, regardless of the threshold specified. Above all, the effect on the amount of risk calculated of the distribution decreased more under MM than under any other parameter estimation method.

The higher the threshold is, the higher the amount of risk tends to be. This may be because the higher the threshold is, the fewer data points there are above the threshold, and consequently, the larger is the impact on the estimates of the high-severity loss data points at the top of the distribution.

When MM is used for parameter estimation, the effect of the threshold on the estimated amount of risk decreases. For example, when a lognormal distribution or a Weibull distribution was assumed, MM yielded similar estimated amounts of risk for different thresholds, whereas differences were greater under MLE and OLS. In addition, for the generalized Pareto distribution, PWM yielded similar estimated amounts of risk for different thresholds. Under OLS, differences were quite large.

5.2.3. Assessment and Discussion of the Results

Using a compound distribution to estimate risk is better than using a single distribution because the choice of distribution and estimation technique has less effect on the quantified amount of risk.

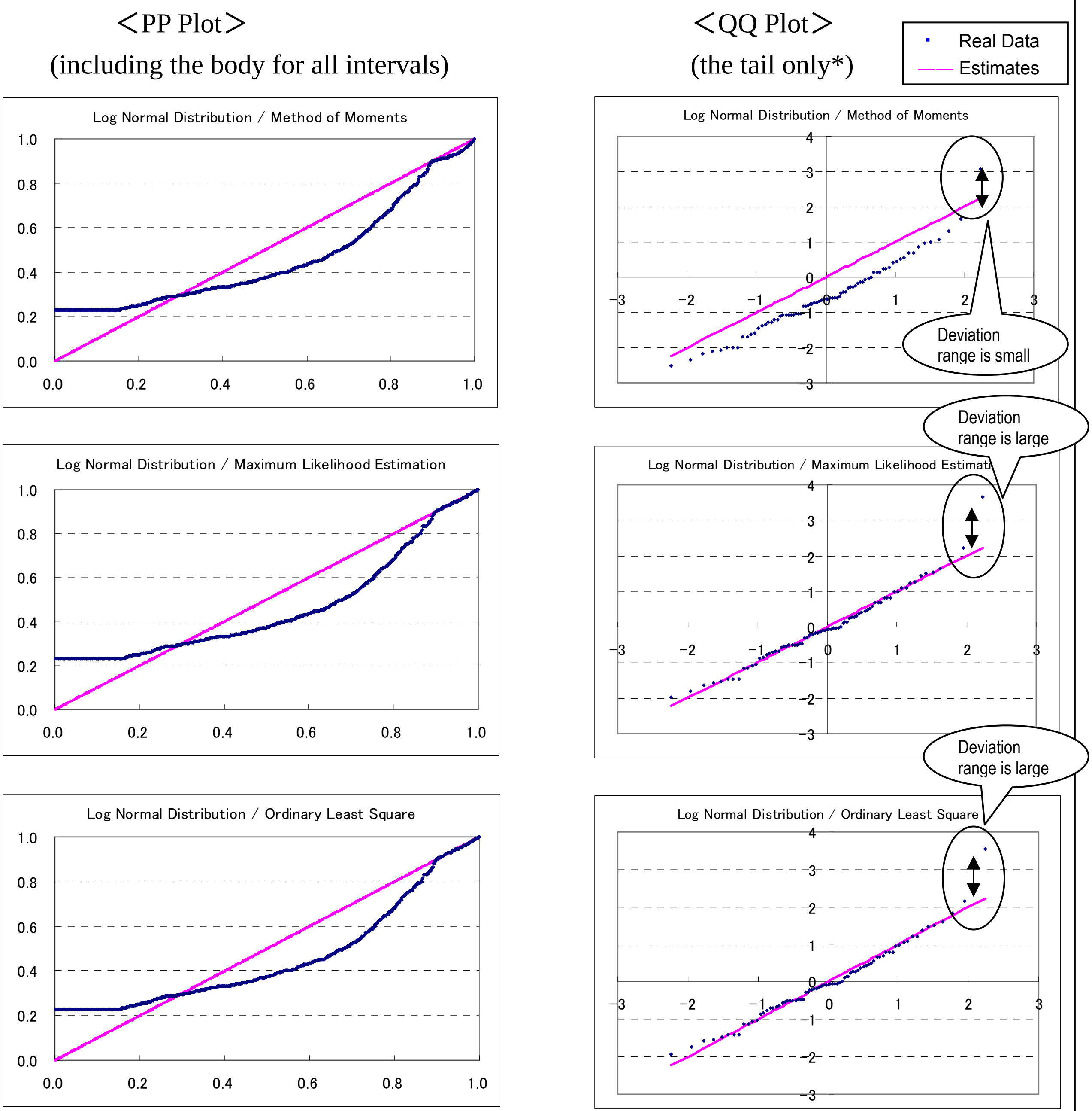
When a lognormal or a Weibull distribution is used for the tail and when MM is used for parameter estimation, the estimated amount of risk is comparable to the benchmark: at the 99% confidence interval, amounts are similar to the benchmark, and at the 99.9% confidence interval, they are approximately 1.5 times the benchmark.

In contrast, caution should be exercised when using a generalized Pareto distribution. Using MLE to estimate a generalized Pareto distribution yielded implausibly large estimated amounts of risk of more than 10,000 times the benchmark (based on the nonparametric method). As when using a single distribution, a generalized Pareto distribution yields very different results under different parameter estimation techniques, even when using a compound distribution.

As we did for the single distribution, for the compound distribution and the parameter estimation method chosen, we use PP and QQ plots, with the threshold at the 90% point, for the lognormal distribution as an example (see [Figure 7] to [Figure 9]).

[Figure 7] Fitness Assessment by PP / QQ Plot When a Compound Distribution is Assumed (Threshold Set at 90% point)

It is demonstrated that the range of deviation in the tail part (which has a greater effect on the result) is smaller when the moment method is used than in cases where the maximum likelihood method or the least square method is used.

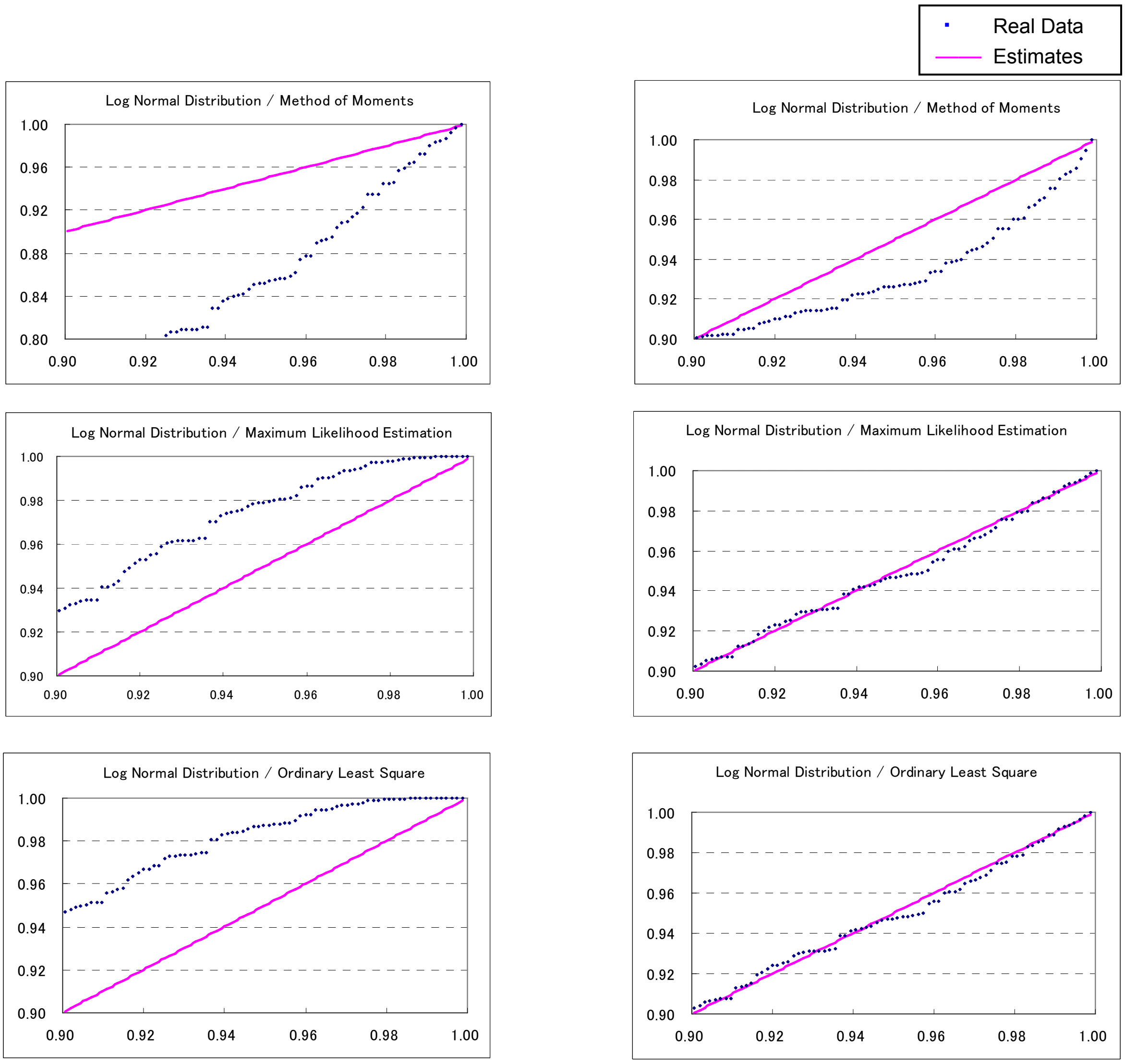


*For the data in excess of the threshold, deviation between the "estimates on the distribution estimated based on the amount in excess of the threshold" and the "amount of the data over the threshold" (also for the QQ plot of the compound distribution shown in Table 9).

[Figure 8] Comparison of Fitness by PP Plot in the Tail (points equal to or over 90%)

If the scope is limited to the portion in excess of 90%, a compound distribution improves the fitness, even when a parameter estimation technique is used, compared to cases where a single distribution is used.

Single distribution (assumes a log-normal distribution) → Compound distribution (assumes a log-normal distribution for both the body and the tail)

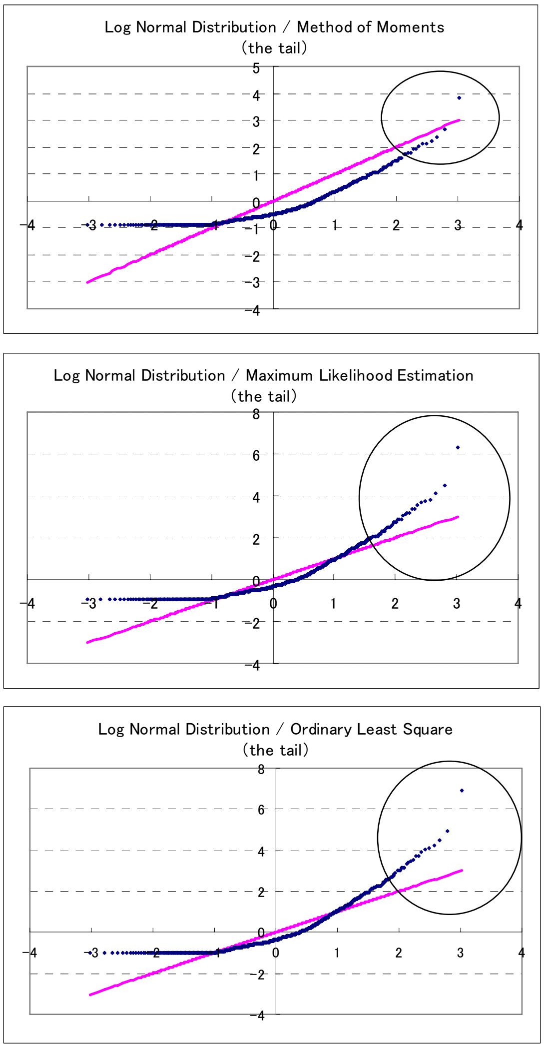


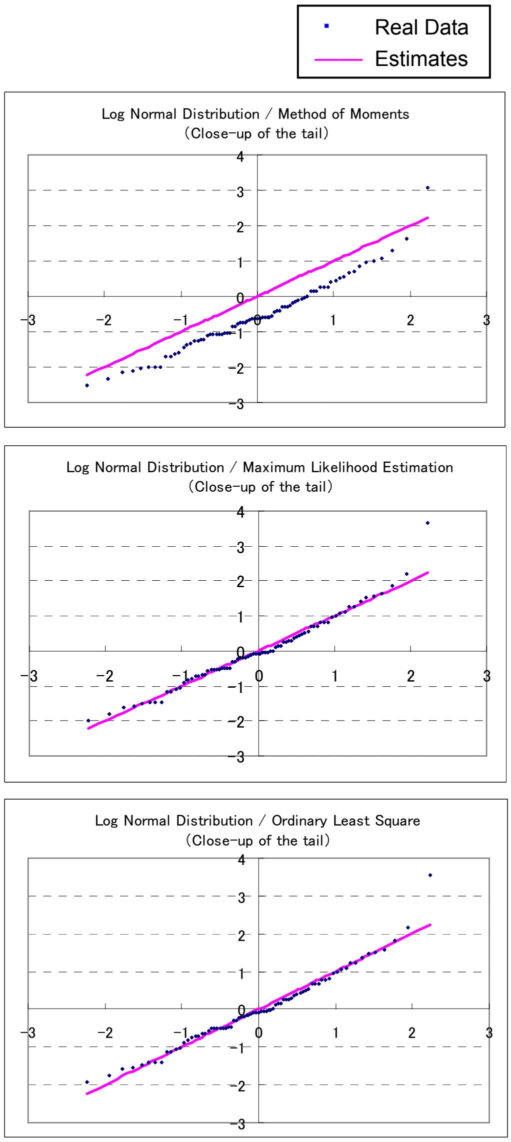
[Figure 9] Comparison of Fitness by QQ Plot in the Tail (points equal to or over 90%)

Additional comparison is made for the degree of fitness by using a QQ plot for cases where a compound distribution is applied. It is clearly shown that the fitness of the right hand side of distributions varies depending on the differences in the parameter estimation methods.

Single distribution
(all intervals including the body)

Compound distribution (the tail only)





1) Application of the Compound Distribution

The range above and including the 90% point suggests that, regardless of the parameter estimation technique used, the estimated compound distribution fits better than does a single distribution. That different parameter estimation techniques produce less variation in the quantification results confirms this.

2) Parameter Estimation Technique

Differences between estimated amounts of risk for different distributions are smaller under MM than under MLE and OLS. There are two possible reasons for this. First, as is the case when a single distribution is used, when a compound distribution is used, the difference between the distribution in the high-severity loss portion above the threshold and the distribution of the data has a smaller range when MM is used than when MLE or OLS is used. Second, the amount of risk is disproportionately affected by the tail of the distribution, relative to the overall distribution.

In addition, because the range of the deviation between these distributions depends on the distribution assumed, the effect is apparently reflected in the differences in the estimated amounts of risk.

As far as our sample data are concerned, to estimate the severity distribution, it is better to use MM, which fits better in the tail, than to use MLE or OLS, both of which fit better in the body.

5.2.4. Issues to be Considered

We discuss two important issues in the context of using a compound distribution: 1) the choice of threshold; and 2) caveats when using a generalized Pareto distribution.

1) Choosing a Threshold

For the calculation method used in this paper, we could not find an effective way of choosing a threshold for a compound distribution based on statistical criteria. Hence, we discuss an alternative method.

One option is to choose a threshold based on the stability of risk quantification. This involves minimizing the impact on risk quantification of changes to the data arising from future loss observations. In other words, it is reasonable to choose the percentile point for the threshold that minimizes the changes in quantified risk arising as new observations on future losses are added to the data set. It should be noted that the value of threshold is sensitive to the addition or deletion of data points, even if the percentile point for the threshold remains constant.

Thus, we tried to choose the percentile point for the threshold that produced relatively small changes in the results when the threshold was changed slightly.

We used this approach because the sample period, and hence the number of data points, was not sufficient. Had this not been the case, it would have been possible to change the sample period and determine an appropriate threshold by assessing changes in the amount of risk.

More precisely, the method we used involved the following. First, we chose a parameter estimation technique that did not produce significant changes in the amount of risk when the threshold was changed. Second, the different risk amounts estimated under different thresholds were used to identify the range of threshold values over which the estimated amounts of risk were relatively invariant. Third, the threshold was chosen from this range.

As an example, we applied the method in the case shown in [Table 10]. In this example, we chose thresholds from a finer scale (90%, 91%, ..., 99%) than those used elsewhere in the paper (90%, 95%, and 99%). Then, we chose thresholds from the range over which differences in the amounts of risk were minimized as much as possible.[30] As a result, the 92% point (at the 99% confidence interval) and the 93% point (at 99.9% confidence interval) were chosen as thresholds.

[30] If the amount of risk at a threshold of p % is C_p, then the difference is defined as:

$$\frac{1}{2}\left(\left|C_p - C_{p-1}\right| + \left|C_{p+1} - C_p\right|\right).$$

[Table 10] Amount of Risk When a Threshold is Set in Stages

Threshold (%)	Confidence Interval			
	99.0%		99.9%	
	Amount of Risk	Difference	Amount of Risk	Difference
90	93.1		297.7	
91	94.5	1.2	307.3	9.1
92	95.5	1.1	298.8	5.8
93	96.7	1.4	295.6	2.0
94	98.4	2.1	296.5	10.2
95	100.9	2.7	316.1	10.3
96	103.8	2.3	317.1	7.3
97	105.5	3.2	303.6	18.5
98	110.2	7.5	327.1	18.3
99	120.4		314.0	

Notes:

1) The amount of risk is the relative value indexed to the value based on the nonparametric method (at 99% confidence), which represents 100.

2) Distribution in the tail is the lognormal.

3) Parameter estimation technique is MM.

4) Threshold: set in increments of 1% within the range 90% to 99%.

5) The number of simulations performed is 100,000.

2) Issues to Consider when Using a Generalized Pareto Distribution

If a generalized Pareto distribution is used for the severity distribution, under MLE and OLS, the estimated amounts of risk may be implausibly large depending on the distribution of the data. For this reason, even if a generalized Pareto distribution is appropriate on theoretical grounds,[31] and on the basis of the suitability of the distribution, implausibly high estimates of risk may be obtained depending on changes in the data.

[31] For example, this is the case when the amount above the threshold, for the portion of data that exceeds the threshold, takes a generalized Pareto distribution when a statistical technique such as the average excess plot is used (see Morimoto [2000]).

6. Conclusion

In this section, we summarize the practical implications and remaining issues regarding the measurement of operational risk.

Using sample data collected from financial institutions, in this paper, we quantified operational risk by using the loss distribution approach (parametric method) with different combinations of distributions and parameter estimation techniques and evaluated the advantages and disadvantages of each measurement technique.

This analysis confirms two points. First, when choosing a measurement method, trying various distributions and different parameter estimation techniques, and then analyzing the differences yields information on how well the assumed severity distributions fit the data. The significant effect on the quantified amount of risk of the parameter estimation technique used suggests that the severity distributions used did not fit the loss data.

Second, for our sample data, a compound distribution fit the data better than did single distributions, which have tails that are heavier than those of the lognormal distribution.

Three issues regarding our analysis remain. First, we were unable to identify a distribution or parameter estimation technique that is generally optimal.

Second, it is difficult in practice to set thresholds for the body and the tail of the distribution in a statistically convincing way. Instead, we proposed a method for setting a threshold based on the stability of the estimated measures of risk when using a compound distribution.

Third, when quantification is based purely on actual loss data, as is the case in this paper, it is difficult in practice to obtain sufficient data for estimating the amount of risk at confidence intervals as high as 99% and 99.9%.[32] For this, data on low-frequency high-severity losses, which are likely to occur approximately every 100 or 1,000 years, are required.[33]

Three issues not discussed in this paper are important.

1) How can one ensure the completeness of the scenario's low-frequency high-severity losses, in particular, when a scenario analysis is introduced as an effective tool for supplementing the low-frequency high-severity loss data, which may not adequately be represented by the actual loss data?

2) How can selection criteria such as event types and business lines for risk-quantification units be established in a rational manner?[34] In addition, how does

[32] To address this issue, as is the practice of financial institutions, it is important to supplement the data with external data and perform scenario analyses.

[33] This can be derived from the relationship between the confidence interval and the range in loss severity distributions that may affect the risk quantification (to what extent high-severity loss events, which rarely occur, should be taken into consideration). (See [Supplementary Discussion 1] for details.) Judging from this result, there is a possibility that the calculations based on the nonparametric method, which are used as a benchmark in this paper, resulted in underestimation.

[34] See footnote 7.

one incorporate the correlation between these units in the amount of risk?[35]

3) How can the stability of the amount of risk be ensured against time-period variations?[36]

Thus, a number of issues remain in the field of operational risk measurement. We intend to continue our analysis and exchange views with experts in this field. We hope that this paper stimulates discussion in this area.

[35] Some financial institutions quantify their operational risks based on a more precise measurement unit such as business lines or event types; from the viewpoint that the characteristics of operational risk they face may be different depending on the business line or event type. However, in reality, no solid model has yet been developed for correlations and dependencies between measurement units.

[36] Because the internal loss data of any financial institution change over time, it is possible that the loss severity distribution assumed by the operational risk measurement model no longer fits the loss data. However, no consensus has yet emerged on how to verify the model's ongoing usability.

[Appendix 1] The Relationship between Confidence Intervals in Risk Measurement and the Range of Loss Data that May Affect the Risk Amount

In Section 1 of this Appendix, we discuss the possibility that the nonparametric method gives understated results, after showing the range of loss data required for calculating the risk amount. First, we show this possibility by deriving a relationship between the confidence interval and the range of the data in the severity distribution that affects the risk amount (to what extent large losses, which rarely occur, should be taken into account in the quantification) based on an example in which a simple nonparametric method is used. Second, we show analytically that a similar relationship exists for the loss distribution approach in general.

In Section 2, the method for estimating operational risk amounts using a closed-form approximation, rather than using a Monte Carlo simulation (hereinafter referred to as simulation), is introduced. In Section 3, the closed-form approximation solution is compared with the simulation results.

1. The Relationship between Confidence Intervals and the Range of Severity Distributions that May Affect the Risk Amount

1.1. Simplified Numerical Examples

In the following examples, the risk amount is calculated for the loss data comprised of two types of loss events (low-frequency high-severity losses and high-frequency low-severity losses) at a number of confidence intervals. That is, the risk amount is calculated through a simulation by using a model with a nonparametric severity distribution for the sample data set that has high-severity loss events with a low frequency that is either once in 100 years or once in 1,000 years.[37]

Structure of the Model:

1) Large Losses that Occur with a Frequency of Once in 100 Years

For the purpose of simplification, all losses other than large losses are assumed to be small and are assumed to occur 100 times in one year. Large losses are assumed to occur once in 100 years. The severities of large and small losses are assumed to be 100,000 and 1, respectively. Because large and small losses respectively occur once and 10,000 times in 100 years, the parameter (the annual average number of loss events) of the loss frequency distribution (Poisson distribution) is:

$$\lambda = 100.01$$

The function *X*, which represents the loss amount, is defined as follows (where p represents the probability, $0 < p < 1$):

$$X(p) = \begin{cases} 1, & 0 \le p < \frac{10000}{10001} \\ 100000, & \frac{10000}{10001} \le p < 1 \end{cases}$$

[37] To verify these results through simulation, it is necessary to perform a sufficient number of trial simulations.

2) Large Losses that Occur at a Frequency of Once in 1,000 Years

Everything is assumed as specified in 1) above except for the frequency (once in 1,000 years) and the severity (1,000,000) of large losses. Since large and small losses respectively occur once and 100,000 times in 1,000 years (as in 1), the parameter (the annual average number of loss events) of the loss frequency distribution (the Poisson distribution) is:

$$\lambda = 100.001$$

The function *X*, which represents the loss amount, is defined as follows (where p represents the probability, $0 < p < 1$) :

$$X(p) = \begin{cases} 1, & 0 \leq p < \frac{100000}{100001} \\ 1000000, & \frac{100000}{100001} \leq p < 1 \end{cases}$$

□

The results:

Given the above assumptions, the risk amount was calculated at four confidence intervals: 99%, 99.1%, 99.9%, and 99.91%. The results are shown in [Table A].

[Table A] Results of Measurement (Comparison of Risk Amounts)

		Occurrence Frequency of Low-frequency Large Losses	
		(i) Once in 100 years	(ii) Once in 1,000 years
Confidence Intervals	99.00%	138	124
	99.10%	100,086	125
	99.90%	100,114	149
	99.91%	100,115	1,000,088

(Number of trials is 100,000[38])

When deriving the risk amount at confidence intervals slightly higher than 99% and 99.9% (99.1% and 99.91%, respectively) from the results above, even large loss events, which occur once in every 100 or 1,000 years, are reflected in the risk amount. This shows that it is necessary to include large loss events that occur once in around every 100 or 1,000 years in the loss data for consideration when the risk amount is being calculated at a confidence interval of 99% or 99.9%. This is because it is assumed that even such large loss events are reflected in the risk amount.

Next, the relationship between the confidence interval and the range of the severity distribution that affects the risk amount, which is suggested by the above simulation, is assessed analytically.

[38] An increased number of trials will give essentially the same result. See the analytical proof given in 1.2 of this Appendix.

1.2. Relationship between the Confidence Interval and the Frequency of Occurrence Required for the Large-loss Data

The relationship between the confidence interval and the range of the frequency of large loss data that may affect the risk amount is shown analytically. Note that the relationship was confirmed numerically in 1.1 with regard to loss events that occur once in every 100 or 1,000 years.

For the purpose of generalization, the question is formulated by assuming a frequency for large losses (with the amount of loss denoted by L_N) of once in N years ($N > 1$) and a frequency for small losses (with the amount of loss being 1 ($1 < L_N$)) of m times per year ($m > 1$).[39] The annual average number of loss events (λ) and the ratio of number of large losses to the total number of losses in N years (R) can be defined by the following equations:

$$\lambda = m + \frac{1}{N}, \qquad R = \frac{1}{N\lambda}, \qquad 0 < R < 1 \quad (\because N\lambda > 1)$$

In addition, the function X, which represents the loss amount, is defined as follows, using R (the ratio of number of large losses to the total number of losses in N years) (where p represents the probability, $0 < p < 1$):

$$X(p) = \begin{cases} 1 \quad \text{(Amount of small losses)}, & 0 \le p < 1 - R \\ L_N \ \text{(Amount of large losses)}, & 1 - R \le p < 1 \end{cases}$$

The probability of observing at least one large loss in the amount of total annual losses, represented by P, can be calculated as follows (where $P_\lambda(x)$ represents the probability distribution of the Poisson distribution):

$$P = \sum_{x=0}^{\infty} P_\lambda(x)\{1 - (1 - R)^x\}, \quad P_\lambda(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

[39] There remains a small possibility that the amount of total annual losses comprising only small losses is equal to or greater than the amount of high-severity losses (P_s). Therefore, the amount of the high-severity loss is set so that this probability (P_s) is sufficiently small. In concrete terms, the probability (P_s) is assessed from the estimated Poisson distribution, and the high-severity loss represented by L_N is determined so that the sum total of this probability and the probability of including at least one high-severity loss represented by (P) (which is estimated as $\frac{1}{N+1} < P < \frac{1}{N}$; see pp. 31–32 for proof) is less than $\frac{1}{N}$. In this way, if the occurrence frequency of the high-severity loss data is once in N years, the probability that the total annual loss is equal to or greater than the high-severity loss, represented by $(P_N = P + P_s)$, can be estimated as $\frac{1}{N+1} < P < P_N < \frac{1}{N}$.

In addition, the amount of the high-severity loss L_N is determined so that the lower is the frequency of occurrence, the higher is the severity. (L_N is an increasing function of N (the reciprocal of the occurrence frequency of high-severity losses)).

The probability P satisfies the following inequality, which depends only on the frequency of occurrence of large losses, and is independent of the annual average number of loss events (λ): (A proof follows.)

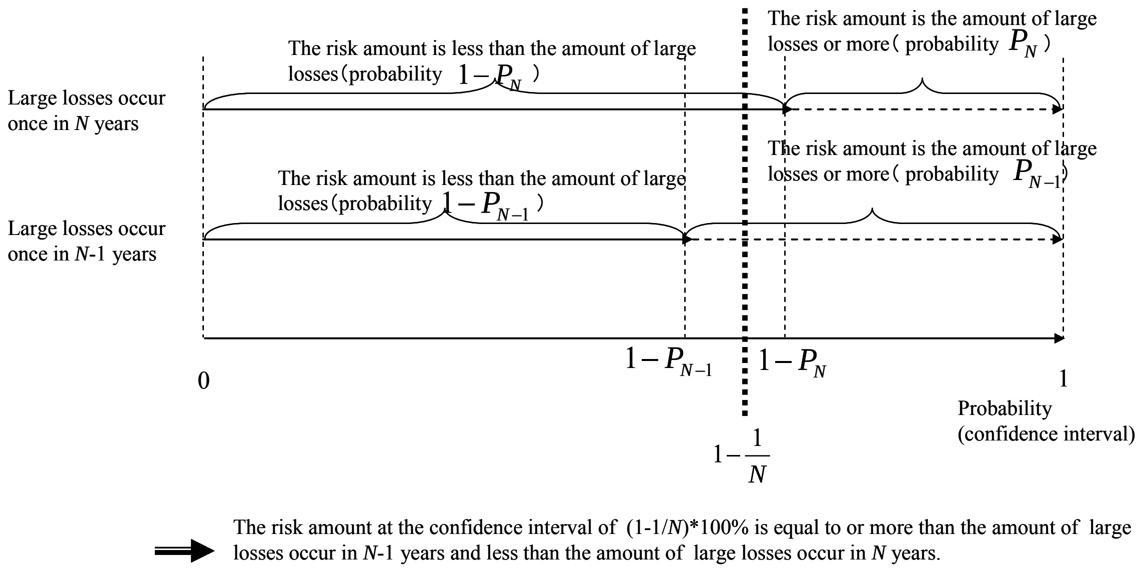
$$\frac{1}{N+1} < P < \frac{1}{N}$$

Accordingly, under the conditions described in footnote 39, if the frequency of occurrence is once in N years for large loss data, the probability that the amount of total annual losses is at least as high as the large loss (P_N) satisfies $\frac{1}{N+1} < P_N < \frac{1}{N}$, independently of the annual average number of loss events. Similarly, if the frequency of the occurrence of a large loss is once in $(N-1)$ years, the probability (P_{N-1}) is $\frac{1}{N} < P_{N-1} < \frac{1}{N-1}$.

The following equation can be derived from the above equations:

$$1-P_{N-1} < 1-\frac{1}{N} < 1-P_N \quad (P_N < \frac{1}{N} < P_{N-1})$$

This shows that while the risk amount estimated at a confidence interval of $(1-\frac{1}{N})\times 100\%$ exceeds the amount of large losses that occur once in $(N-1)$ years, this will not exceed the amount of large losses that occur once in N years. In other words, the risk amount estimated at such a confidence interval is between the large loss that occurs once in $(N-1)$ years and the large loss that occurs once in N years (see the figure below).



Therefore, this suggests the following three points be considered in estimating the risk amount at a confidence interval of $(1-\frac{1}{N})\times 100\%$. First, large loss events that occur at a frequency of once in $(N-1)$ years should be considered. Second, the large loss events which occur at a frequency of less than once in N years do not affect the result. Third, large loss events that occur at a frequency between once in $(N-1)$ years and once in N years should be taken into account as there is a possibility that such losses affect the result (thus, along with the second point, it is sufficient to take into consideration large loss events that occur at a frequency of no less than once in N years).[40]

This suggests, for example, that to estimate the risk amount at confidence intervals of 99% and 99.9%, it is at least necessary to take into consideration large loss events that occur at a frequency of no less than once in 99 or 999 years. It also suggests that large loss events that occur at a frequency of no more than once in 100 or 1,000 years have no impact on the risk amount when estimating that amount at confidence intervals of 99% and 99.9%. This is consistent with the measurement results shown in [Table A].

Financial institutions in general are unlikely to accumulate data on the occurrence of such large loss events. Therefore, when calculating the risk amount by using the nonparametric method on loss data, and in particular when calculating the risk amount at a confidence interval of 99.9% or higher, it is possible that such a value will be understated.

Proof of the inequality $\frac{1}{N+1} < P < \frac{1}{N}$ $(for\ 1 < N)$:

$$P = \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!}\{1-(1-R)^x\} = \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} - \sum_{x=0}^{\infty}(1-R)^x = 1-e^{-\frac{1}{N}}$$

(i) Proof of $P < \frac{1}{N}$

If we set $H(N) = \dfrac{1}{N} - P = \dfrac{1}{N} - (1-e^{-\frac{1}{N}}) = \dfrac{1}{N} - 1 + e^{-\frac{1}{N}}$

Then, $H(N) > 0$ is proved as follows:

$$H'(N) = -\frac{1}{N^2} + \frac{1}{N^2}e^{-\frac{1}{N}} = -\frac{1}{N^2}\left(1-\frac{1}{e^{\frac{1}{N}}}\right)$$

As $e^{\frac{1}{N}} > 1$ $(for\ 1 < N)$, it follows that $H'(N) < 0$.

Then, as $H(1) = 1-1+e^{-1} = \frac{1}{e} > 0,\ \ \lim_{N\to\infty} H(N) = 0\ (\because \lim_{N\to\infty} e^{-\frac{1}{N}} = 1)$.

Therefore, $H(N) > 0$ $(for\ 1 < N)$.

$$\therefore P = \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!}\{1-(1-R)^x\} < \frac{1}{N}.$$

(ii) Proof of $\frac{1}{N+1} < P$

[40] This may be true if the VaR is adopted as an indicator of the risk amount. However, if some other indicator is used (such as an expectation shortfall), low-frequency high-severity loss events may have a substantial impact on the amount of risk.

If we set $G(N) = P - \dfrac{1}{N+1} = (1 - e^{-\frac{1}{N}}) - \dfrac{1}{N+1} = -\dfrac{1}{N+1} + 1 - e^{-\frac{1}{N}}$.

Then, $G(N) > 0$ is proved as follows:

$$G'(N) = \frac{1}{(N+1)^2} - \frac{1}{N^2} e^{-\frac{1}{N}} = \frac{1}{(N+1)^2} \underline{\left\{1 - (1 + \frac{1}{N})^2 e^{-\frac{1}{N}}\right\}}$$

The underlined part of the right-hand side of the first equation is defined as:

$g(N) = 1 - (1 + \frac{1}{N})^2 e^{-\frac{1}{N}}$, and let $t = \frac{1}{N}$, $(where\ 0 < t < 1)$. Then:

$$g(t) = 1 - (1+t)^2 e^{-t}$$

$$\therefore g'(t) = -e^{-t}(1 - t^2) < 0 \quad (\because 0 < t < 1)$$

Therefore, $G'(N) < 0 \quad (for\ 1 < N)$.

Then, as $G(1) = \frac{1}{2} - \frac{1}{e} > 0$, $\lim_{N \to \infty} G(N) = 0$, $G(N) > 0 \quad (for\ 1 < N)$.

$$\therefore \frac{1}{N+1} < \sum_{x=0}^{\infty} \frac{e^{-\lambda} \lambda^x}{x!} \{1 - (1-R)^x\} = P$$

From (i) and (ii) above, we obtain

$$\tfrac{1}{N+1} < P < \tfrac{1}{N} \quad (1 < N) \qquad \blacksquare$$

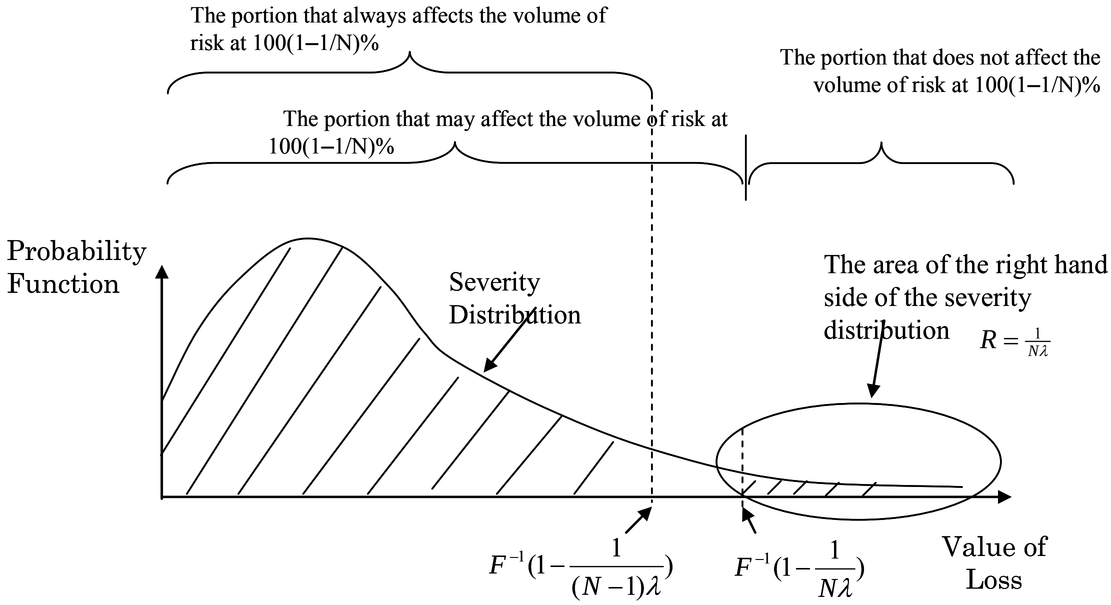
1.3. The Relationship between the Confidence Interval and the Range of the Severity Distribution that May Affect the Risk Amount

The relationship between the confidence interval and the frequency of occurrence required for the large-loss data is not limited to particular cases such as the examples given in 1.1 and 1.2 above (for the nonparametric method, which assumes only two types of loss value); it is applicable to any loss distribution approach in general. In other words, to measure the risk amount at a confidence interval of $(1 - \frac{1}{N}) \times 100\%$ on an annual basis, it is at least necessary to make precise estimates up to the $(1 - \frac{1}{(N-1)\lambda}) \times 100\%$ point from the bottom of the severity distribution. Points down to the $\frac{1}{N\lambda} \times 100\%$ point or above from the top have no impact on the risk amount (see [Figure A]).[41, 42]

[41] As in the cases of 1.1 and 1.2, in which only two types of severity were assumed, this relationship holds true when the probability is sufficiently small that the amount of total annual losses that include at least one high-severity loss event (the right-hand side of the $\frac{1}{N\lambda}$ part from the top of the severity distribution in [Figure A]) is smaller than the amount of total annual losses that include no high-severity loss.

[42] However, in the case of the parametric method, even if the range of the severity distribution that affects the risk amount is specified, the data required to undertake parameter estimation of the severity distribution are not limited to this range. Rather, it is necessary to estimate the parameters by including high-severity loss data.

[Figure A] The Range of the Severity Distribution That Affects the Risk Amount (Graphic Image)



2. Operational Risk Measurement Based on a Closed-Form Approximation

In the previous subsection, we identified the range of severity distribution that affects the calculation of the risk amount derived from the relationship between confidence intervals and the occurrence frequency of losses. Based on a similar concept, Böcker and Klüppelberg [2005] developed a method of approximating the risk amount at a particular point on the loss distribution (the value of one loss event). According to them, if the severity distribution function is subexponential,[43] the risk amount $VaR(\alpha)$ at a confidence level of 100α%, given an average total annual number of events of λ and using the severity distribution function F, can be approximated by:

$$VaR(\alpha) = F^{-1}(1-\frac{1-\alpha}{\lambda})$$

If we set $N = \frac{1}{1-\alpha}$, this equation can be written as:

---

[43] The distribution function is subexponential when the distribution function $F(\cdot)$ and its n-fold convolution $F(\cdot)^{n*}$ satisfy the following:

$$\lim_{x\to\infty}\frac{1-F(x)^{n*}}{1-F(x)} = n \quad ,n \geq 2 \quad here, 1-F(x)^{n*} = \Pr\left(\sum_{i=1}^{n} X_i > x\right)$$

This means that if x is large enough, the probability that the value of total losses (the amount of risk) exceeds x (the probability that the total amount of losses extracted independently n times from the same distribution exceeds x) can be approximated by the probability that the amount of loss extracted once from such a severity distribution exceeds x, multiplied by n. See Embrechts *et al.* [1997] for details.

The lognormal distribution, the Weibull distribution, and the generalized Pareto distribution (if the shape parameter ξ is positive), which are adopted as parametric severity distributions in this paper, are all subexponential.

$$VaR(\alpha) = F^{-1}(1 - \frac{1}{N\lambda})$$

This means that when N represents the reciprocal value of the occurrence frequency of large losses (in other words, if a large loss occurs once in N years as defined in Part 1 of Appendix 1), the risk amount can be approximated as the amount of loss at the percentile point that ranks $\frac{1}{N\lambda}$ from the top of the severity distribution. For example, to approximate the risk amount at confidence intervals of 99% and 99.9%, a single statistic is needed: the amount of the loss event that occurs at a frequency of once in every 100 or 1,000 years when a nonparametric distribution is assumed.

In summary, Böcker and Klüppelberg argue that when the confidence interval is sufficiently high and when the severity distribution is subexponential, the upper limit of the range of the severity distribution that affects the risk amount, as shown in 1 of Appendix 1, can be deemed to be the approximate value of the risk amount.

This suggests that, when estimating the risk amount through simulation, when the confidence interval is high enough, and when the severity distribution is subexponential, the tail (at around the abovementioned upper limit, in particular) has a large impact on the estimation results and plays an important role. (Other parts of the distribution have little impact.)

3. Comparison of the Approximated VaR (by Closed-form Approximation) and the Simulated VaR (by Monte Carlo Simulation)

In this subsection, we assess the approximation accuracy of the risk amount obtained from the closed-form approximation solution proposed by Böcker and Klüppelberg [2005] by comparing it with our simulation results (see [Table B] on the next page).

First, in general, when the same distribution is assumed, the accuracy of the closed-form approximation tends to improve as the risk amount calculated increases (i.e., as the rate of deviation decreases). This may be because, when the tail is relatively heavy in the estimated distribution, the approximation for the VaR will be more accurate; by contrast, if the tail is relatively thin, the approximation for the VaR will be less accurate.[44]

---

[44] To use the closed-form approximation solution, as a prerequisite, the following approximate expression, shown in footnote 43, must exist for the total value of the losses calculated (x):

$$\frac{1 - F(x)^{n*}}{1 - F(x)} = n \quad , n \geq 2$$

As explained in footnote 43, all parametric distribution used in this paper for estimation of the severity distribution are subexponential. Although it is certain that when $x \to \infty$, the above approximate expression is appropriate, it may not be appropriate in the vicinity of the total amount of losses represented by x .

This means that the probability that the value of the total amount of loss (the volume of risk) exceeds x (the probability that the total value of the losses extracted independently n times from the same loss distribution exceeds x) may no longer be approximated by the probability that the amount of loss extracted once from the severity distribution exceeds x , multiplied by n . For example, it may be impossible to ignore the probability that the total amount of loss (the volume of risk) exceeds x , although the amount of all n loss events extracted from the severity distribution is no more than x .

Furthermore, in all cases, the closed-form approximation has a smaller value than that resulting from simulation. Thus, the closed-form approximation may be understated compared to the risk amount that is calculated from the estimated severity distribution.

As explained above, when the tail in the estimated distribution is not very heavy, the accuracy of the closed-form approximation may decline. Moreover, the approximation may understate the risk amount. Accordingly, when using the closed-form approximation for operational risk quantification, it is necessary to check the accuracy of the approximation.

[Table B] Comparison of Closed-Form Approximation Solutions and Simulation Results

Distribution	Parameter Estimation Technique	Closed-Form Approximation Solution[45]		Simulation Result (see [Table 3] in the main text)		Rate of Deviation[46]		Distribution Parameter	
		Confidence Intervals						Scale	Shape
		99.0%	99.9%	99.0%	99.9%	99.0%	99.9%		
Lognormal Distribution	MM	62.0	243.8	74.9	272.6	-17.3%	-10.6%	2.17	2.47
	MLE	1.3	3.2	2.5	4.4	-47.4%	-27.0%	1.48	1.61
	OLS	0.8	1.8	1.8	2.8	-55.1%	-35.4%	1.48	1.47
Weibull Distribution	MM	92.2	321.5	105.8	351.7	-12.9%	-8.6%	0.70	0.18
	MLE	1.5	2.5	3.9	5.0	-62.1%	-49.4%	10.85	0.43
	OLS	0.3	0.4	1.6	1.8	-81.5%	-76.8%	10.90	0.63
Generalized Pareto Distribution	PWM	24.6	232.1	26.8	255.4	-8.5%	-9.1%	4.57	0.98
	MLE	50.1	630.4	54.1	686.1	-7.5%	-8.1%	3.44	1.10
	OLS	1.1	4.5	1.8	5.5	-37.3%	-17.8%	3.84	0.60

This is more likely if the total amount of loss (the volume of risk) is relatively low (i.e., if there is relatively little tail heaviness in the estimated severity distribution). Therefore, if there is relatively little tail heaviness, the closed-form approximation solution may be less accurate.

[45] These represent the scaled values based on the volume of risk at a confidence interval of 99%, with the values obtained by using the nonparametric method representing 100.

[46] This is the rate of deviation of the closed-form approximation solution from the simulation result, which is defined as:

$$\frac{\text{Closed-form approximation solution—Simulation result}}{\text{Simulation result}}$$

[Appendix 2] Technical Terms Used and Remaining Issues

1. Distribution Used for Severity and their Characteristics

	Characteristics / Shape	Probability Density Function	Parameter
Lognormal Distribution	The natural logarithm of the loss value x follows a normal distribution. (The probability distribution is $x=e^{Y}$, where Y follows a normal distribution: $N(\mu,\sigma^{2})$.) Defined within the range of $x>0$, the tail becomes heavier as x increases.	$f_{LN}(x)=\frac{1}{\sqrt{2\pi}\sigma x}\exp\left(-\frac{(\log x-\mu)^{2}}{2\sigma^{2}}\right)$, $x>0$	μ : Average of the natural logarithm of x (scale parameter) σ : Standard deviation of the natural logarithm of x (shape parameter)[47]
Weibull Distribution	Defined within the range of $x>0$; if the parameter p is smaller than 1, the tail becomes heavier as x increases.	$f_{w}(x)=\frac{p}{\theta}\left(\frac{x}{\theta}\right)^{p-1}\exp\left(-\left(\frac{x}{\theta}\right)^{p}\right)$, $x>0,\quad p,\theta>0$	p : Shape parameter θ : Scale parameter
GPD: Generalized Pareto Distribution	Defined within the range of $x>0$, the tail becomes heavier as x increases.	$f_{GPD}(x)=\frac{1}{\beta}(1+\frac{\xi}{\beta}x)^{-\frac{1}{\xi}-1}$, $x\geq 0,\quad \xi\neq 0,\quad \beta>0$ $f_{GPD}(x)=\frac{1}{\beta}\exp(-\frac{x}{\beta})$, $x\geq 0,\quad \xi=0,\quad \beta>0$	ξ : Shape parameter (the larger is ξ, the heavier is the tail) β : Scale parameter

The generalized Pareto distribution has the following characteristics, depending on the value of ξ. First, if $\xi<0$, there exists some L such that the probability of the amount of loss exceeding L is zero. Second, if $\xi>\frac{1}{r}$, no r th-order moment exists for the distribution. In particular, if $\xi>1$, the distribution has more tail heaviness with no average value (or first-order moment).

[47] This represents a parameter that is held constant through the use of scale conversion. (All data are multiplied by a fixed number other than zero.)

2. Parameter Estimation Techniques and their Characteristics

	Concept	Methodology	Characteristics
Method of Moments (MM)	Simultaneous equations are developed for a moment (e.g., a parameter relating to an average or variance) and the solution is calculated to obtain an estimate of the parameter.	Based on loss data, the first- and second-order moments are calculated as follows: $M_1 = \frac{1}{N}\sum_{i=1}^{N} x_i, M_2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2$ (1) Then the formulas for M_1, M_2 expressed using the parameters for the assumed distribution are prepared (They vary depending on the distribution assumed) (2). Then simultaneous equations of (1) and (2) are solved to obtain the parameters.	In calculating the moment, this method tends to emphasize the portion with relatively large values of x (low-frequency, high-severity part, which represents the tail of the distribution).
Probability-weight Moment Method (PWM)	A weight is assigned to each data point in order to calculate the moment; the larger the value of the data point, the less weight is assigned. This method is applied to a distribution with quite a heavy tail for which no moment such as an average or variance may exist, depending on the value of the parameter.	Based on loss data, the transformed moment is calculated as: $w_r = E[X(1-F(X))^r] \quad (r = 0,1,\cdots)$ (1) $F(X)$: distribution function. Then the formulas for w_r expressed using the parameters for the assumed distribution are prepared (These vary depending on the distribution assumed) (2). Then simultaneous equations of (1) and (2) are solved to obtain the parameters.	This method is used instead of MM for a distribution with quite a heavy tail, such as the generalized Pareto distribution.
Maximum Likelihood Estimation method (MLE)	The parameter is set so that the likelihood (expressed as the product of the occurrence probability density of the loss data that represent the sample) is maximized. Specifically: 1) The loss data are assigned to the probability density function of the assumed distribution to obtain a function of the parameter. 2) The product of the values (function) obtained in 1) above is calculated to obtain the likelihood function. 3) Determines the parameter so that the value of the likelihood function is maximized.	Based on loss data, the log likelihood function is calculated as: $\Lambda = \log\prod_{i=1}^{N} f(x_i,\theta_1,...,\theta_K) = \sum_{i=1}^{N}\log(f(x_i,\theta_1,...,\theta_K))$ and $\frac{\partial\Lambda}{\partial\theta_j} = 0, \quad j = 1,...,K$ is solved to derive the parameter θ_j. For the log normal distribution, two parameters θ_1, θ_2 represent the sample average and the sample variance, respectively, of the log value of the loss data.	In the calculation of the log likelihood function, this method tends to estimate parameters that emphasize the fitness of the high-frequency low-severity part, which has more data. The maximum likelihood estimator is usually the asymptotically efficient estimator (the variance of the estimator is minimized as the number of observations $n \to \infty$).

Ordinary Least Square method (OLS)	The sum of the squares of the differences between the loss data and the estimate (the point on the estimated distribution function) is computed and the value of the parameter is determined so that the value calculated represents the minimum value.	For the combination of data converted so that the loss data and the estimate have a linear relationship (y_i, x_i), the parameter is determined so that $\sum\{y_i-(ax_i+b)\}^2$ is minimized. For example, if a log normal distribution is assumed for the severity distribution, y_i is the monotonically increasing sequence of the log value of the loss data, and x_i is the monotonically increasing sequence when the log value of the loss data exactly matches a standard normal distribution.	As with MLE, there is a tendency to emphasize the fitness in the high-frequency low-severity portion of the distribution, which has more data..

3. Visual Validation Methods for Fitness of Distribution (PP and QQ Plots)

Name	Concept	Points when used for verification of the fitness of the loss distribution.
PP Plot (percentile–percentile plot)	For the assumed distribution and the distribution of the loss data, the status of deviation on the probability point of each datum (which represents the position in terms of percentage from the lowest value in the range 0–1) is shown; x-axis: the estimate on the estimated distribution; y-axis: the value based on the loss data.	The closer are the points to the 45-degree line, the better is the fit of the model. It is also possible to visually identify any under- or overestimation of the loss data (any points above the 45-degree line understate the loss data, while those below the line overstate it).
QQ Plot (quantile–quantile plot)	For the assumed distribution and the distribution of the loss data, the status of deviation of each loss value is shown with modifications and adjustments added as appropriate to facilitate visualization; x-axis: the estimate on the estimated distribution; y-axis: the value based on the loss data.	In a QQ plot, it is easier to comprehend the deviation between the loss data and the estimates in the high-severity portion visually, as the axes are modified based on the severity. On the other hand, in a PP plot, it is difficult to visually distinguish the deviation in the large loss portion.

4. Quantitative Methods for Verification of Fitness of Severity Distributions and How they are Appraised

Name	Concept	Method used in this paper	Reasons for difficulties in using the method for evaluating the goodness of fit of severity distributions
Kolmogorov–Smirnov Statistics (K–S Statistics)	At every point of the loss data, the (absolute value of the) difference between (a) the distribution of the loss data and (b) the estimated distribution is calculated and the maximum value is taken as the value of the statistic.	$K = \max_{1 \le i \le N} \lvert S_N(x_i) - F(x_i) \rvert$ K : K–S statistic value; $S_N(x_i)$: The distribution function of the loss data distribution, etc (the percentage point when the data is arranged in ascending order); $F(x_i)$: Estimated distribution function for N , the number of data points; x_i : Value of each datum.	The deviation in the high-frequency low-severity portion is emphasized (because the distribution function of the low-frequency high-severity portion is close to the value of unity for both distributions, and the deviation does not become too large). For this reason, it is difficult to assess appropriately the deviation in the low-frequency high-severity portion, which is expected to have a large effect on the risk amount.

Anderson–Darling Statistics (A–D Statistics)	A revised version of the K–S statistic. The statistic's value is the sum of the square of the differences between the distribution function of the loss data and the distribution function of the estimated distribution multiplied by the weighting function $\psi(x)$.	$A^2 = \int_{-\infty}^{+\infty} \left\| S_N(x) - F(x) \right\|^2 \psi(x) dF(x),$ $\psi(x) = \frac{N}{F(x)(1-F(x))}$ A^2 : A–D Statistic value N : Number of data points	It is generally recognized that the statistic value is sensitive to the deviation in the tails on both the left and right sides of the distribution (because high weights are placed on the tails at both ends of the distribution by the weighting function $\psi(x)$). However, as in the case of the K–S statistic's value, it is difficult to appropriately assess the deviation in the low-frequency, high-severity portion, because the deviation for the tail on the left, i.e., for the high-frequency low-severity portion, is adopted as the value of the statistic (the distribution function of the low-frequency high-severity portion is close to unity for both distributions, the deviation does not become too large).

References

Böcker, C. and C. Klüppelberg, "Operational VaR: A Closed-Form Approximation," *Risk*, 18(12) pp. 90–93, 2005.

Chapelle, A., Y. Crama, G. Hübner and J. Peters, "Basel II and Operational Risk: Implications for Risk Measurement and Management in the Financial Sector," National Bank of Belgium, 2004 (http://www.nbb.be/doc/ts/publications/WP/WP51En.pdf).

de Fontnouvelle, P., V. DeJesus-Rueff, J. Jordan and E. Rosengren, "Capital and Risk: New Evidence on Implications of Large Operational Losses," Federal Reserve Bank of Boston, 2003 (http://www.bos.frb.org/economic/wp/wp2003/wp035.pdf).

de Fontnouvelle, P., J. Jordan and E. Rosengren, "Implications of Alternative Operational Risk Modeling Techniques," NBER Working Paper No. W11103, 2004 (http://www.algorithmics.com/solutions/opvantage/docs/OpRiskModelingTechniques.pdf).

Dutta, K. and J. Perry, "A Tale of Tails: An Empirical Analysis of Loss Distribution Models for Estimating Operational Risk Capital," Federal Reserve Bank of Boston, 2006 (http://www.bos.frb.org/economic/wp/wp2006/wp0613.pdf).

Embrechts, P., C. Klüppelberg and T. Mikosch, *Modeling Extremal Events for Insurance and Finance*, Berlin, Germany: Springer–Verlag, 1997.

Hoaglin, David C., "Summarizing Shape Numerically: The g-and-h Distributions," in David C. Hoaglin, Frederick Mosteller and John W. Tukey (eds.), *Exploring Data Tables, Trends, and Shapes*, New York, NY: John Wiley & Sons, Inc., 1985, pp. 461–513.

Klugman, S., H. Panjer and G. Willmot, *Loss Models*: *From Data to Decisions* (2nd ed.), New York, NY: John Wiley & Sons, Inc., 2004.

Mitsubishi Trust & Banking Corporation Operational Risk Study Group, "Operational Risk no subete (The Practice of Operational Risk Management)", Toyo Keizai Inc., 2002 (in Japanese).

Morimoto, Yuji, "Kinyu to Hoken no Yugo ni tuite (Fusion of Finance and Insurance)," Kinyu kenkyu, Vol. 19 (S–1), pp. 289–342, Institute for Monetary and Economic Studies, Bank of Japan, April 2000 (in Japanese).

Moscadelli, M., "The Modeling of Operational Risk: Experience with the Analysis of the Data Collected by the Basel Committee," The Bank of Italy, 2004 (http://www.bancaditalia.it/ricerca/consultazioni/temidi/td04/td517/en_tema_517.pdf).

Study Group on the Advancement of Operational Risk Management, "Discussions on Further Advancing Operational Risk Management (1) – Summary of Discussions of the "Study Group on the Advancement of Operational Risk Management," June 2006 (http://www.boj.or.jp/type/release/zuiji_new/fsc0606a.pdf).