Risk Management Group Conference
on
Leading Edge Issues in Operational Risk Measurement

Integrating External Data into the OR Measurement Approach

Sumitomo Mitsui Financial Group

Operational Loss data

- Concern

  -Insufficient internal historical loss data, especially for tail events.

- Issues

1. Fail to capture potential risks (i.e. tail events)
2. Unprecedented large loss amount has huge impact on marginal capital (i.e. lack of robustness)

Neither parametric nor non-parametric approach can deal with the above issues well.

	Description	Strength	Weakness
Parametric	Choosing a distribution and estimating its parameters	Generates a potentially fat tail.	-Lacks robustness -Less powerful when parametric assumptions are not met.
Non-parametric	Sampling from the empirical distribution	No parametric assumption about the distributions and parameters.	Does not generate a potentially fat tail.

Smoothed Bootstrap Methodology

- An intermediate solution between parametric distribution and non-parametric bootstrapping.

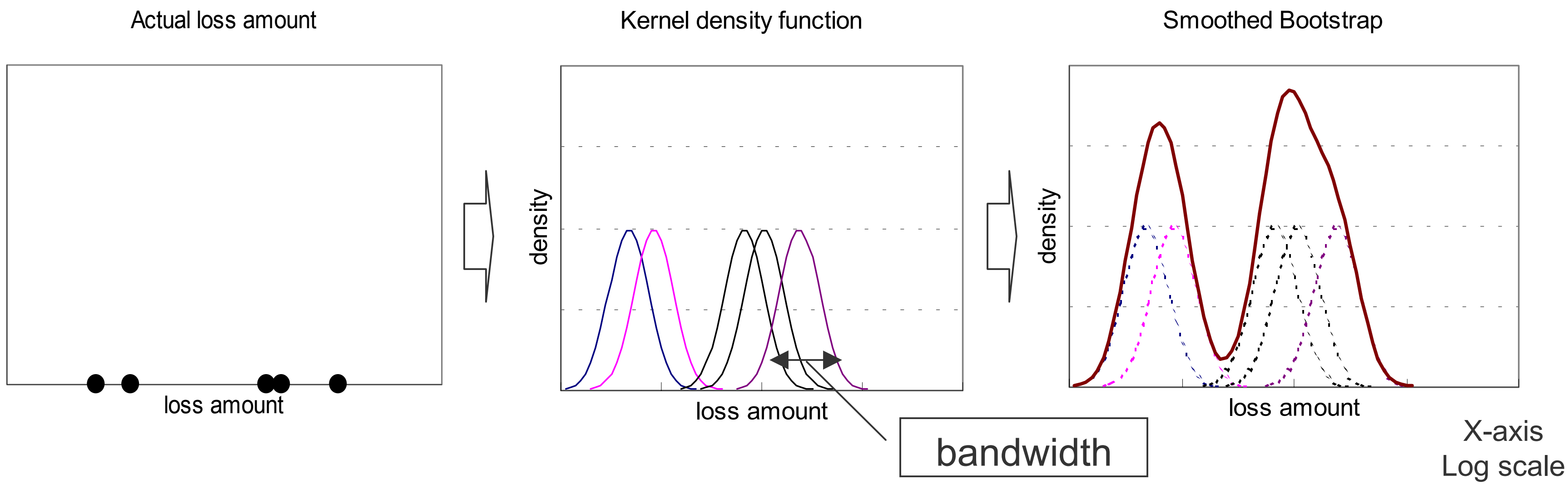
	Description	Strength	Weakness
Parametric	Choosing a distribution and estimating its parameters	Generates a potentially fat tail.	-Lacks robustness -Less powerful when parametric assumptions are not met.
Non-parametric	Sampling from the empirical distribution	No parametric assumption about the distributions and parameters.	Does not generate a potentially fat tail.

Smoothed bootstrap	***Generates a potentially fat tail without any assumptions about distributions.***

- Easy mixture of internal and external data (or scenarios).

Smoothed Bootstrap Methodology

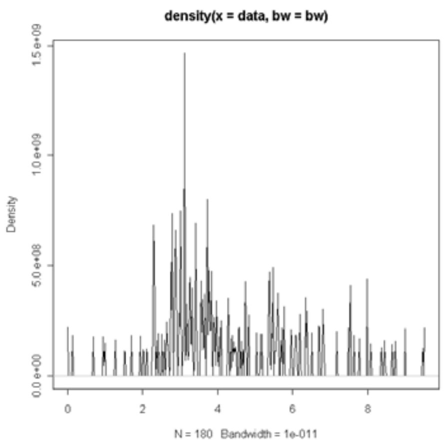
- Methodology of smoothing and sampling



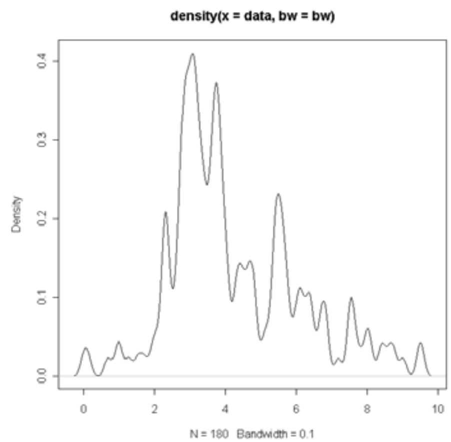
  - Instead of re-sampling directly from the empirical distribution, smooth it first then the smoothed distribution is used to generate new samples. (Monte Carlo method)

Smoothed Bootstrap Methodology

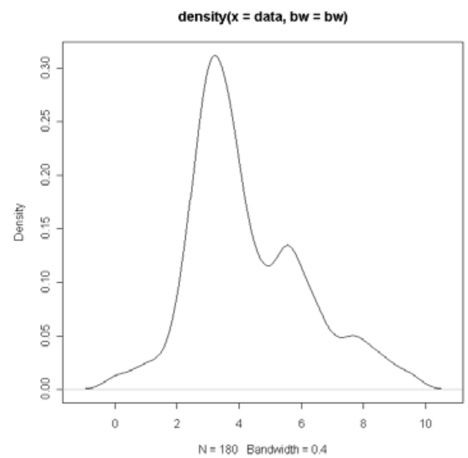
- Tail behavior of the smoothed distribution
  – Sample observations N=180



Bandwidth=0



Bandwidth=0.1



Bandwidth=0.4

X-axis
Log scale

Distribution →

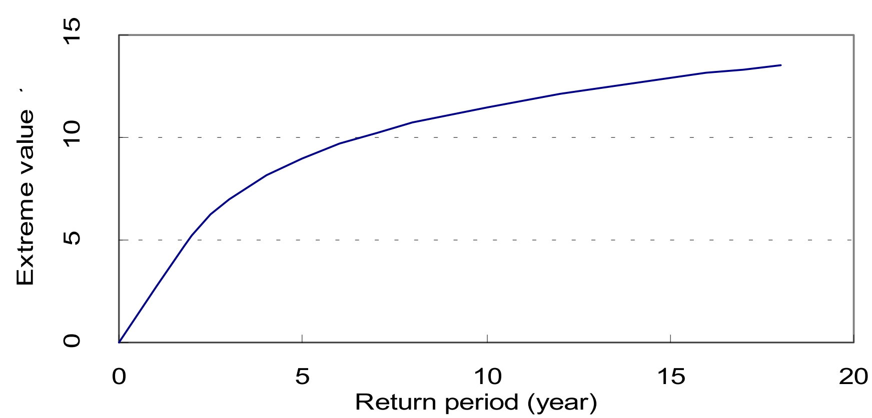
(Smoother fatter tail)

  – The larger the bandwidth, the fatter the tail of the distribution.

- Once the frequency and severity of the tail events are given by internal data, external data or scenarios, the bandwidth can be determined.

Extreme value during a return period

- The frequency and severity of the tail events are described as "Extreme value X during the N-year return period"
  - X is a threshold that is exceeded once per N-year return period on average.
- We get X and N by applying Gumbel distribution.



Process of smoothed bootstrap Application

- How to collect enough data to characterize the maximum value distribution

Case ◊ : Use External Data

Case ◊ : Use Scenario Analysis

Application [Case 1: External Data]

■ Case 1: External Data

Select maximum loss amount for each year.

Bank C

Bank B

Bank A

year	maximum loss amount
****	****
1995	10,000
1996	500,000
1997	200,000
1998	90,000
1999	120,000
2000	350,000
2001	1,000,000
2002	100,000
****	****

Scaling by exposure (size) indicator [revenue, assets etc.]

Determine . and å of the maximum loss amount

Determine parameters a and b of the Gumbel distribution

Determine X(N)

Determine the bandwidth of the Smoothed Bootstrap

Application [Case 2: Scenario Analysis]

■ Case 2: Scenario Analysis

Scenario data can be collected by using self-assessment including interviews, workshops and questionnaires.

<<Scenario output>>

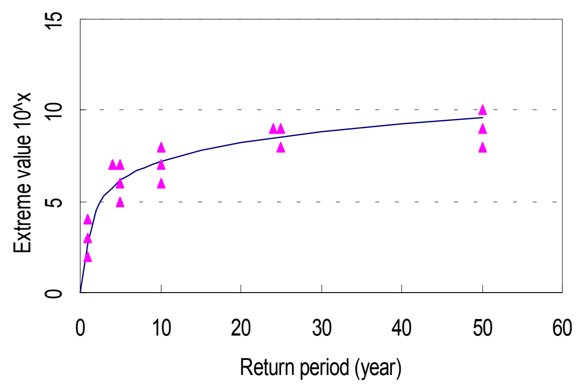
Scenario type: ***********

Respondent	Frequency	Severity
a	2	3
b	3	2
c	1	4
d	3	3
e	2	4
	4	2
	1	5

Frequency:
1 : once every 50 years
2 : once every 25 years
3 : once every 10 years
4 : once every 5 years
5 : once a year

Severity:
1 : <100K
2 : >100K and <1mil
3 : >1mil and <10mil
4 : >10mil and <100mil
5 : >100mil

Regression analysis and curve fitting



Plot :
Each assessment

Determine X(N)

Determine the bandwidth of the Smoothed Bootstrap

Integration of internal and external data (scenarios)

- X(N) determined by *only* internal data.
  - Does not capture severe tail loss events.
  - Too sensitive to large loss amounts.

- X(N) determined by *only* external data (or scenarios).
  - Less sensitive to internal large loss amounts.
    - ✓ Required capital does not increase despite the occurrence of a large internal loss event.
    - ✓ Required capital does not decrease despite the improvement of the internal control.

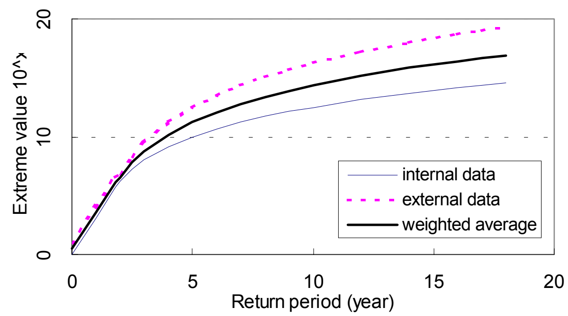
⇩

- Determine X(N) by *both* internal and external data (or scenarios).
  - Capture potential risks.
  - Reflect internal control factors.
  - Sensitive, but not too sensitive, to large loss amounts.

Integration of internal and external data (scenarios)

- Combined distribution of X $_{internal}$(N) and X $_{external}$(N)

$$Y \equiv \omega X_{internal}(N) + (1-\omega) X_{external}(N)$$



✓ Combined distribution above is again the Gumbel distribution with the parameters = $\omega\mu_{internal} + (1-\omega)\mu_{external}$ and
= $\omega\sigma_{internal} + (1-\omega)\sigma_{external}$

✓ is determined for each event type.

✓ Assess weighting of internal and external data using factor analysis or similar method.

Test of Robustness

■ Sample calculation (How much marginal impact does the maximum loss amount have ?)

Sample observation
- Observation period : [n+1] • [n+5] year

	Number of loss events	Total loss amount US$ mil.	Maximum loss amount US$ mil.	Parametric model Unexpected loss US$ mil. (1y, 99.9%)	Smoothed Bootstrap Unexpected loss US$ mil. (1y, 99.9%)
Event type 1	58	11	5	189	150
Event type 2	5,012	39	13	212	223
Event type 3	212	110	30	1,231	970
Event type 4	•••	•••	•••	•••	•••
Event type 5	•••	•••	•••	•••	•••

Next year

Sample observation
- Observation period : [n+2] • [n+6] year

	Number of loss events	Total loss amount US$ mil.	Maximum loss amount US$ mil.	Parametric model Unexpected loss US$ mil. (1y, 99.9%)		Smoothed Bootstrap Unexpected loss US$ mil. (1y, 99.9%)	
Event type 1	60	17	**6**	354	**• 165**	174	**• 24**
Event type 2	4,895	55	**20**	312	**• 100**	343	**• 120**
Event type 3	223	322	**200**	3,803	**• 2,572**	1,520	**• 550**
Event type 4	•••	•••	•••	•••		•••	
Event type 5	•••	•••	•••	•••		•••	

- Internal data
  - Historical data for 5 years
- External data
  - Fixed
- Weighting of internal and external data
  - Internal data : 0.5
  - External data : 0.5

= Summary =
Smoothed Bootstrap Methodology

- Smoothed Bootstrap Methodology is a desirable solution especially for small data sets.
  - Captures potentially severe tail loss events.
  - Reflects both
    - Internal loss data for calibrating the main body of the severity distribution, and
    - External loss data or scenarios for calibrating the tail of the severity distribution.
  - Robust about outlier data and sensitive to internal control.

Appendix

Smoothed Bootstrap Methodology

- Define

$$\hat{f}(x) = \frac{1}{b}\sum_{j=1}^{n} K\left(\frac{x - x_j}{b}\right)$$

where

» K: Kernel density function

» x_j: Observations   j=1 to n

» b: Bandwidth

- Suppose
  – K : Normal distribution with mean x_j and standard deviation b
  – x_j : Log-normal of loss amount (j=1 to n)

Smoothed Bootstrap Methodology

- Bandwidth can be determined by the frequency and severity of the tail event, which is described as
  "Extreme value X during the N-year return period"

  – X is a threshold that is exceeded once per N-year return period on average.

  – X and N are related by the following equation :

$$\int_X^{\infty} \frac{1}{b} \sum_{j=1}^{n} K\left(\frac{x - x_j}{b}\right) dx = \frac{1}{Accumulated\ loss\ number\ for\ N\ years} \quad [*1]$$

Probability of loss amount exceeding X

Once every N years

- Once X and N are given by the internal data, external data or scenarios, the bandwidth b can be determined by solving the equation [*1] for b.

Distribution of the maximum value

- Distribution of the maximum value
  - Given independent identically distributed (i.i.d.) random variables $X_1, X_2, \ldots, X_n$ with an underlying distribution function F.
  - How to obtain the distribution fitting into maximum $M_n = \max\{ X_1, X_2, \ldots, X_n \}$.

    e.g. distribution of the maximum loss amount selected for each year.

- Extreme Value Theory (EVT)

  Distribution of the Maximum Value is of one of the following three types:

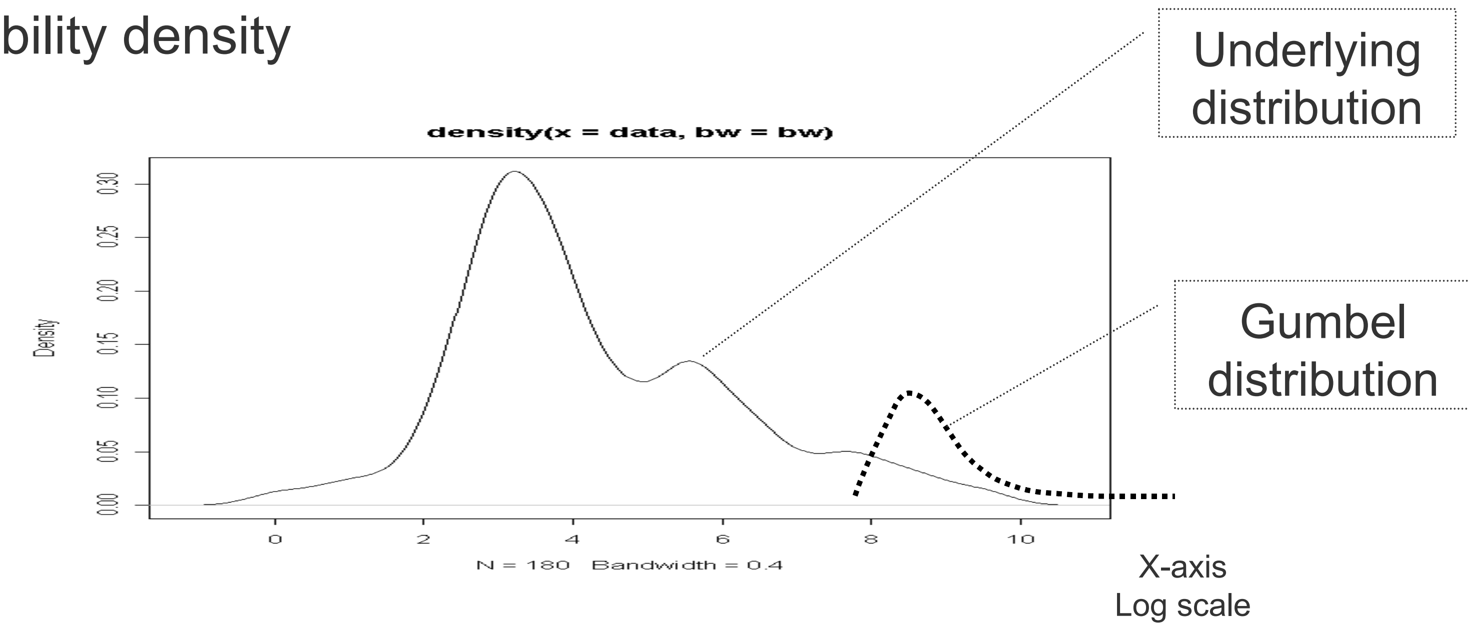
  - ✓ Type I (Gumbel-type) : The case where F belongs to a class of ***medium*-tailed** distributions, including normal, exponential, gamma and log-normal distributions.
  - ✓ Type II (Frechet-type) : The case where F belongs to a class of ***heavy*-tailed** distributions, including Pareto, log-gamma, Cauchy etc.
  - ✓ Type III (Weibull-type) : The case where F belongs to a class of ***short*-tailed** distributions, characterized by a finite right endpoint, including uniform and beta distributions.

Gumbel distribution

- Extreme value Type I distribution
  - The case where underlying distribution F belongs to a class of ***medium*-tailed** distributions, including normal, exponential, gamma and log-normal distributions.
- Cumulative distribution function

$$G(x,a,b) = \exp(-\exp(-a(x-b)))$$

- Probability density



Extreme value during a return period

- We get X(N) ("Extreme value X during the N-year return period") by applying Gumbel distribution as follows :

  - Frequency is the inverse of the return period N.

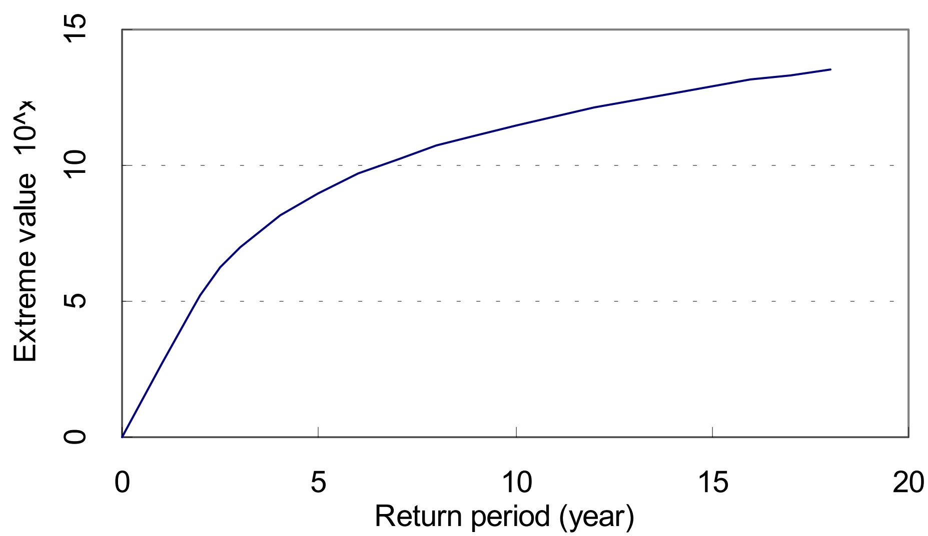
  - X is the value of the upper 1/Nth percentile of the Gumbel cumulative distribution.

$$G(X,a,b) = \exp(-\exp(-a(X-b))) = 1 - \frac{1}{N} \qquad [*2]$$

Extreme value during a return period

- Solving equation [*2] for X, we get

$$X(N) = b + \frac{1}{a}\left[-\ln\left\{-\ln\left(1-\frac{1}{N}\right)\right\}\right] \qquad [*3]$$



Parameter estimation for the Gumbel distribution

- Parameter estimation
  - Estimators with the moment method for the Gumbel distribution are

$$a = \frac{\pi}{\sigma\sqrt{6}} \qquad b = \mu - \frac{\sigma\sqrt{6}}{\pi}\gamma \qquad [*4]$$

where

- μ and σ are the mean and standard deviation of the maximum loss amount observed for a given period, respectively.
- γ is Euler's number (constant 0.5772).

Process of smoothed bootstrap

- Observe maximum loss amount for each year.
- Calculate ¸ and å of the maximum loss amount.
- Determine parameters a and b of the Gumbel distribution.

$$a = \frac{\pi}{\sigma\sqrt{6}} \qquad b = \mu - \frac{\sigma\sqrt{6}}{\pi}\gamma \qquad [*4]$$

- Determine X(N).

$$X(N) = b + \frac{1}{a}\left[-\ln\left\{-\ln\left(1-\frac{1}{N}\right)\right\}\right] \qquad [*3]$$

- Determine the bandwidth of the Smoothed Bootstrap.

$$\int_X^{\infty} \frac{1}{b}\sum_{j=1}^{n} K\left(\frac{x-x_j}{b}\right)dx = \frac{1}{Accumulated\ loss\ number\ for\ N\ years} \qquad [*1]$$