An Analysis of Japanese Banks' Loss Data: Possible Use of External Loss Data

Tsuyoshi Nagafuji
Basel2 Implementation Office
General Coordination Division
Supervisory Bureau
Financial Service Agency
(On secondment from Bank of Japan)

This presentation and material outlines preliminary results of the presenter's research done while he was at the BOJ. It does not necessarily express established views or policies of the FSA or the BOJ.

1. Objectives of the presentation

This presentation intends to:

- present an example utilizing external data:
  - ✓ to show a possible use of external data for operational risk management;
  - ✓ to present materials for discussion in this workshop.

  The analysis was done by the BOJ team on the data collected in the 2007 Operational Risk Data Collection Exercise (conducted February 2007; published August 2007).

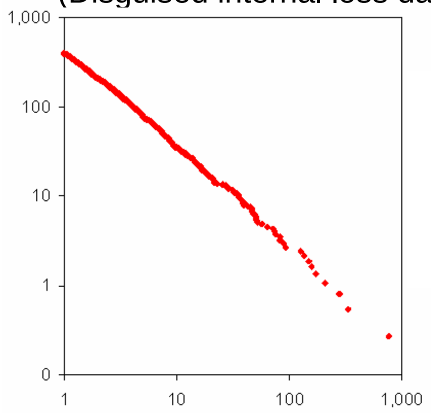
  Please note that the methods and results of this analysis are not intended to be applied by private financial institutions for their risk management, although some of the methods and results can be useful for such institutions.

- further investigate common characteristics of the operational risk losses of Japanese banks that are available in the 2007 Operational Risk Data Collection Exercise.

2. Methods
(1) Introduction (“Double logarithmic graph”)

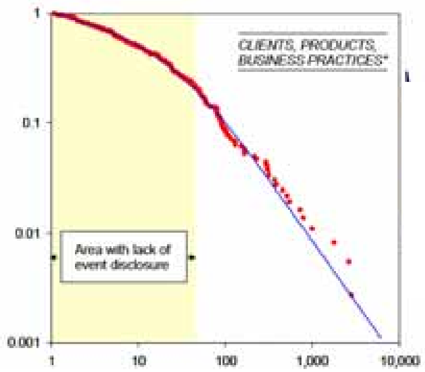
- Loss data from some banks around the globe show a similar pattern in the “Double logarithmic graph”, in which the log values of (cumulative) frequency and severity have a linear relationship, indicating the power law relationship.

<Ex1> Citi
(Disguised internal loss data)



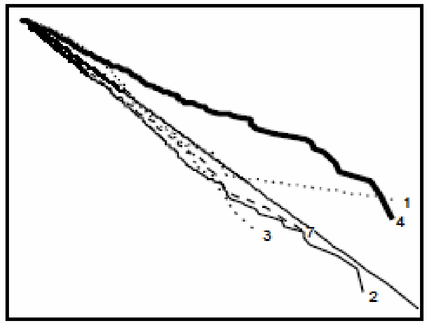
Source: Dekoker and Newberry (2005)

<Ex2> Published loss data



Source: SAS®, Op Risk Global Data, Dekoker and Newberry (2005)

<Ex3> US LDCE 2004



Source: de Fontnouvelle et al. (2004)

- Similar patterns were observed in the losses collected in the 2007 Operational Risk Data Collection Exercise.
- I will discuss some possible interpretations of the so-called “Double logarithmic graph.”

(2) Outline of the data

- This analysis uses the internal loss data collected in the 2007 Operational Risk Data Collection Exercise.
  - ✓ This exercise requested information on operational risk data (internal loss data and scenario data) from 14 banks (including bank holding companies) that use or plan to use internal loss data for the calculation of operational risk capital.

- Important points about the data set are as follows.
  - ✓ The data set was collected in February 2007.
  - ✓ Fourteen Japanese banks submitted their operational risk internal loss data.
  - ✓ Collected information includes the loss amount (net/gross), the date of the loss (date of occurrence and date of discovery), associated business lines and types of occurrences. Detailed descriptions of losses were not collected.
  - ✓ The data period submitted by the banks differed from bank to bank, ranging from more than seven years to less than a year.

    →To ensure comparable annual averages, we used 'stable' losses, which excluded the losses that had been collected before putting their current loss data collection systems in place.

(An example)

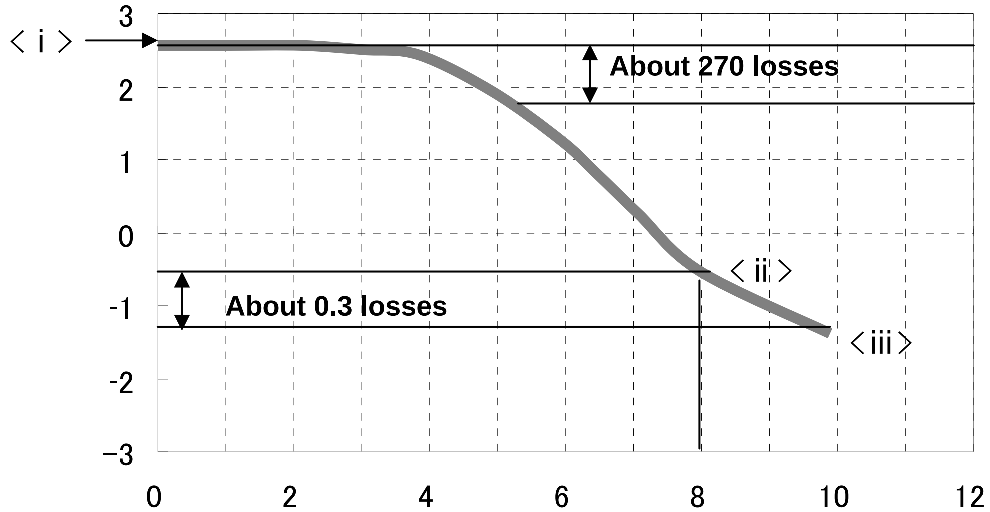
					Stable losses						
Fiscal half-year	01-1	01-2	02-1	02-2	**03-1**	**03-2**	**04-1**	**04-2**	**05-1**	**05-2**	**06-1**
Number of losses	2	3	4	8	**110**	**82**	**119**	**134**	**123**	**111**	**135**

  - ✓ Net losses equal to 1 yen or more were used.
  - ✓ Losses related to credit risk (0.6% of total number of losses, 5.3% of total amount of losses) were used.

(3) How to read the graph (A simulated example)

All losses are plotted on the graph.
X: Severity (yen, common logarithm) (not scaled)
Y: Frequency (number of losses greater than or equal to each loss)
(scaled: per year, per 10 trillion yen in total assets)

Frequency
(number of losses greater than or equal to a particular loss)



Severity

<i>: Number of losses greater than or equal to Yen 10^{0} yen (1 yen) equals $10^{2.5}$ (about 320).

<ii>: Number of losses greater than or equal to 10^{8} yen (100,000,000 yen) equals $10^{-0.5}$ (about 0.3).

<iii>: The maximum loss is about 10^{10} yen (10,000,000,000 yen); occurrence $10^{-1.3}$ (about 0.05).

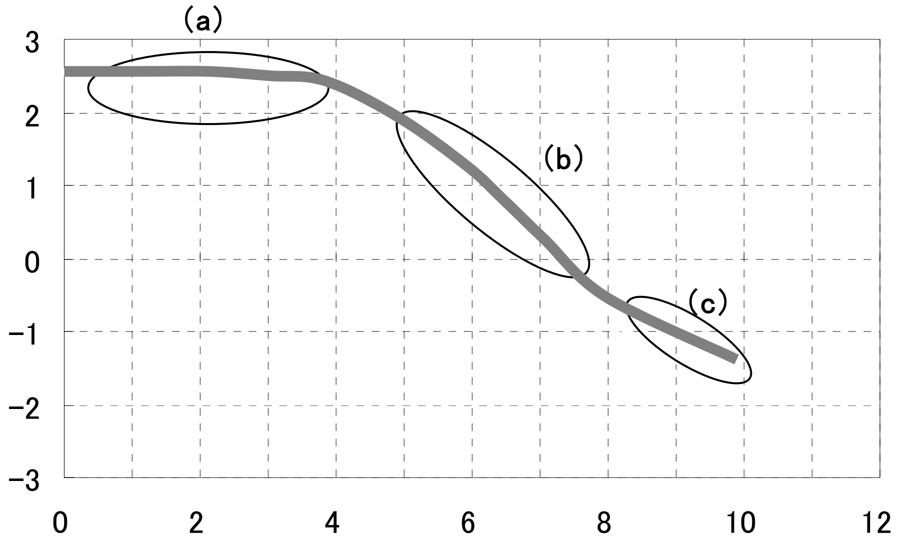
For example:0.05=1(Total number of losses) /10 (years of data collection) / scaling factor (e.g. 2 for 20 trillion yen bank)

3. Analyses

(1) Analyses by banks (from a simulated example)

- Shapes and locations of the curves are strikingly similar among the participating banks.

Frequency (number of losses greater than or equal to a particular loss)



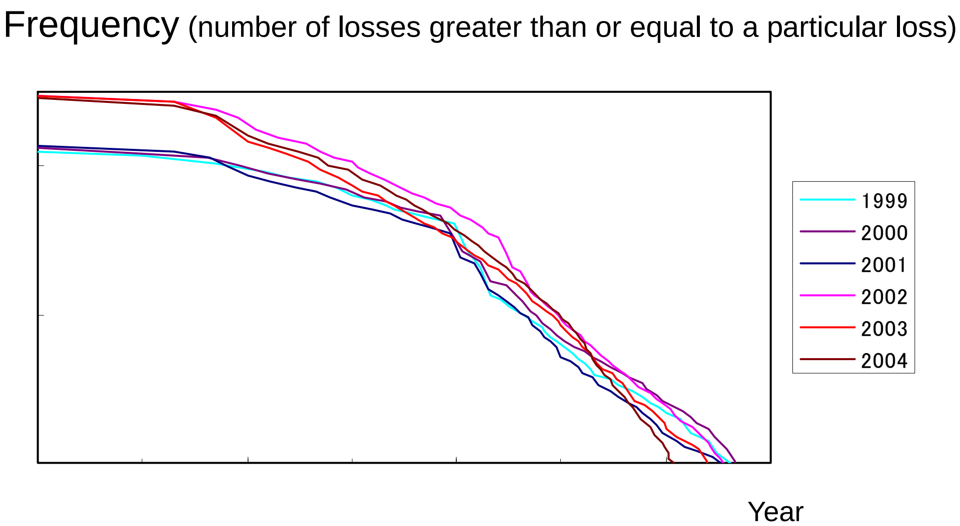
Severity

<a> Gentler slopes—some are almost horizontal—below 10^4–10^5 yen.
→ This can be because of the nature of operational losses,
→ or this may suggest less complete loss data collection.

<b> In the middle, the curve becomes almost linear with a slope of –1.

<c> Gentler slopes between 10^7–10^8 yen.
→ This may be because of the paucity of the loss data; a longer data collection period may result in a steeper slope,
→ or this may suggest difficulty in containing “infrequent/large losses.”

■ Curves appear to be stable over time. Major shifts in the curves show changes in the methods and breadth of the data collection.

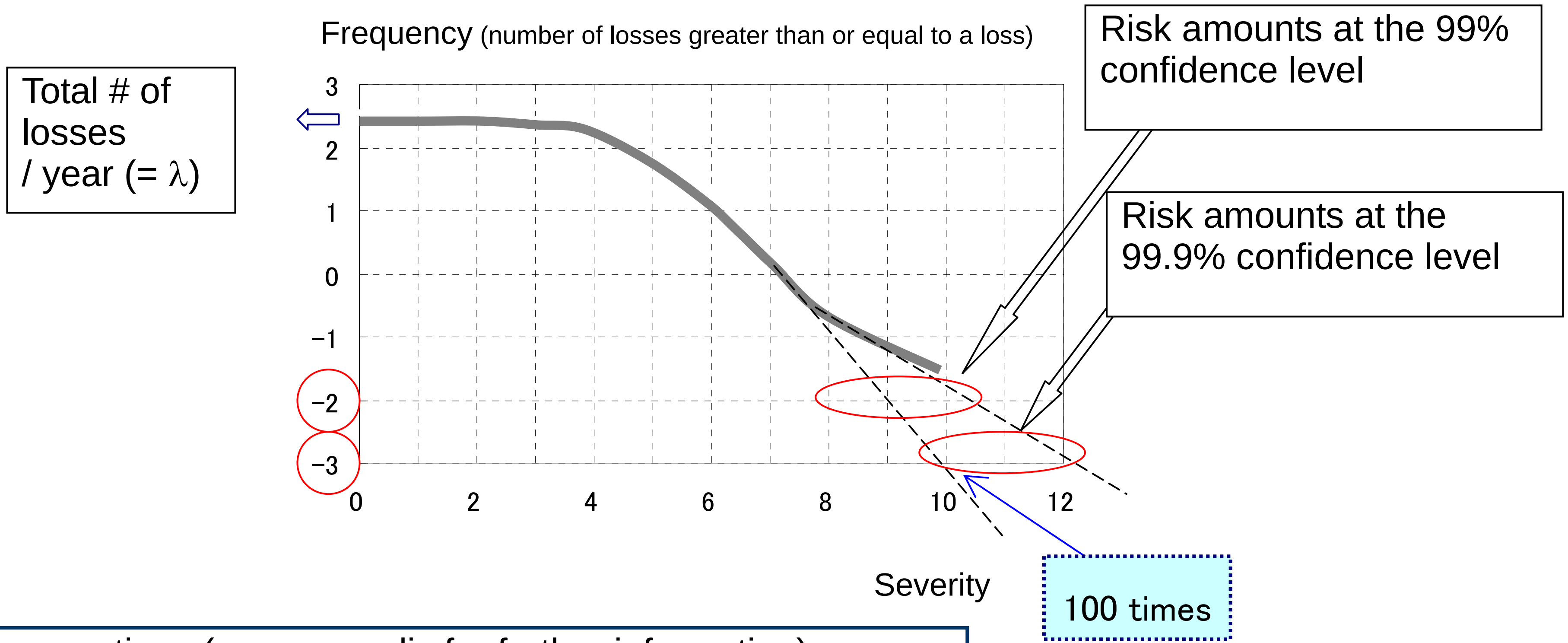


The graph is realistic but hypothetical; it is based on multiple financial institutions.
→Upward shift is observed between 2001 and 2002.

- ✓ Changes in the methods and breadth of the data collection caused the shift.
- ✓ Upward shift is observed in the small loss area (<a> in the previous graph), while no significant shift is observed in the middle (<b> in the previous graph) and the tail (<c> in the previous graph).
- ✓ With the upward shift, some banks' curves show leftward movements of turning points between the small loss area and the middle area.

■ Risk amounts can be inferred from the graph.

Intersection of the extrapolation of the graph and the line parallel to the X-axis gives the approximate risk amount at the confidence level measured by the Y-axis.



Assumptions (see appendix for further information):
(a) “Single loss approximation” gives a good approximation.
(b) Straight line extrapolation is meaningful.

(2) Losses by Basel category

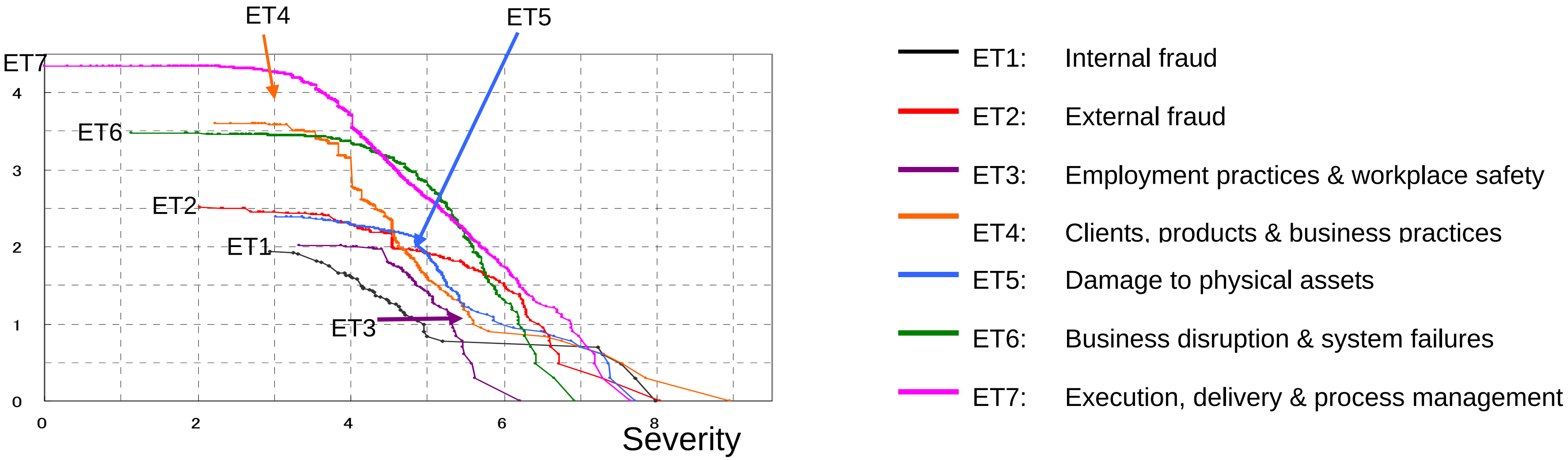
(Event types)

- Loss data from several banks are aggregated by the event types.
  - Graph 1: The same as the previous graphs
  - Graph 2: Modified from the previous graphs:
    - ✓ plots only the losses equal to or greater than 10^5 (100,000) yen, which seem to be more complete than the smaller ones;
    - ✓ plots the fraction of losses exceeding each threshold.

- Observations:
  - ✓ Turning points between small loss area and medium loss area differ by event type(graph1).
    →Comprehensiveness of the data appears to differ by the event type.
  - ✓ Gaps at the loss amount of 10,000 yen are observed(graph1).
  - ✓ Risk profiles appear different between event types, although many event types do not have enough losses to derive any concrete conclusions(graph2).

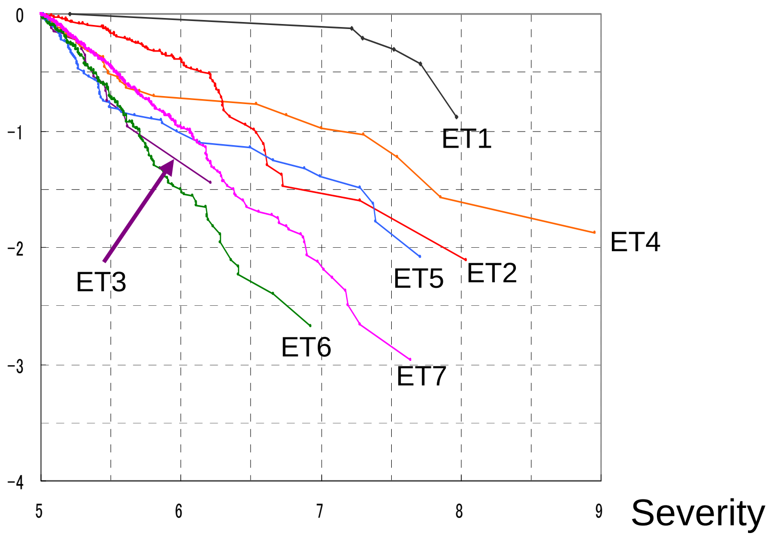
Graph 1

Frequency (number of losses greater than or equal to each threshold)



Graph 2

Frequency (fraction of losses exceeding each threshold, only for losses equal to or greater than 100,000 yen)



(Business lines)

- Loss data from several banks are aggregated by the event types.

  Graph 1: The same as the previous graphs

  Graph 2: Modified from the previous graphs:

  - ✓ plots only the losses equal to or greater than 10^5 (100,000) yen, which seem to be more complete than the smaller ones;
  - ✓ plots the fraction of losses exceeding each threshold.

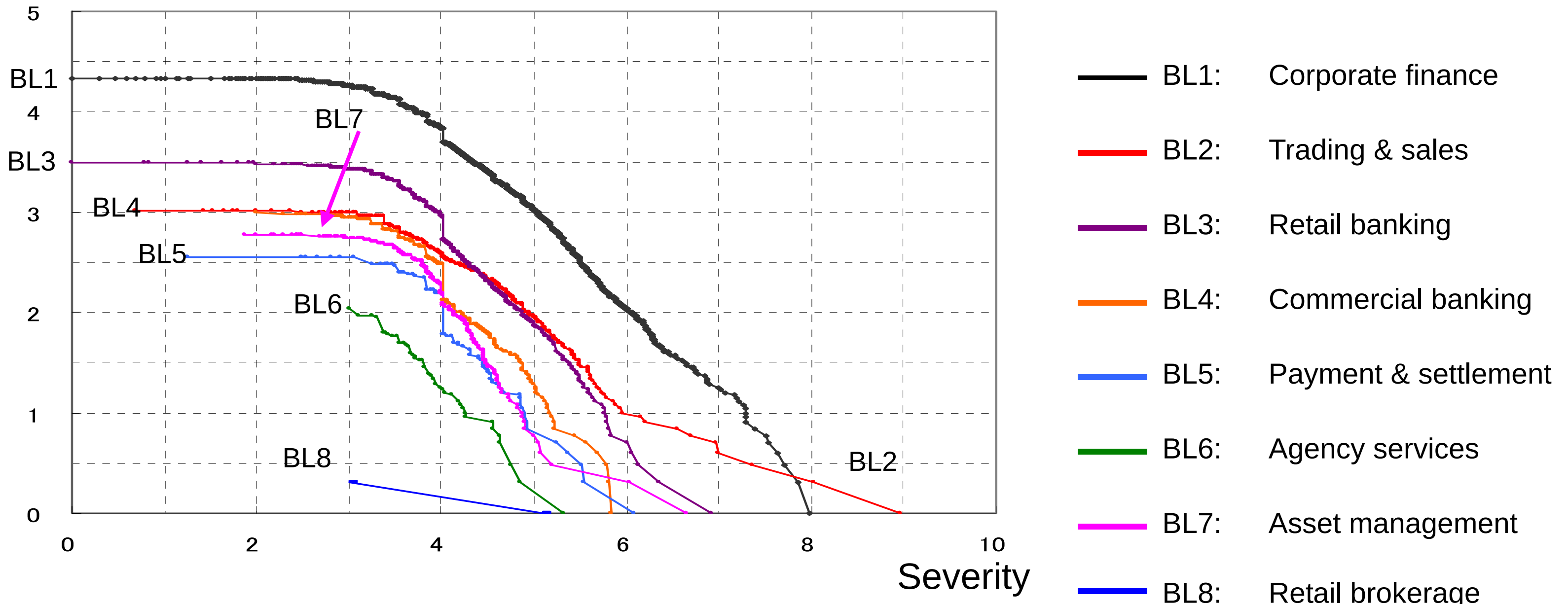
- Observations:
  - ✓ Turning points between small loss area and medium loss area differ by business line (graph 1).

    →Comprehensiveness of the data appears to differ by business line.
  - ✓ Curves look more similar to the event type curves (graph2).

    →This may be one of the reasons for Japanese banks' preference for measuring operational risks by event type rather than by business line.
  - ✓ BL2 and BL3 appear to be riskier, although many business lines do not have adequate losses to derive any concrete conclusions (graph2).

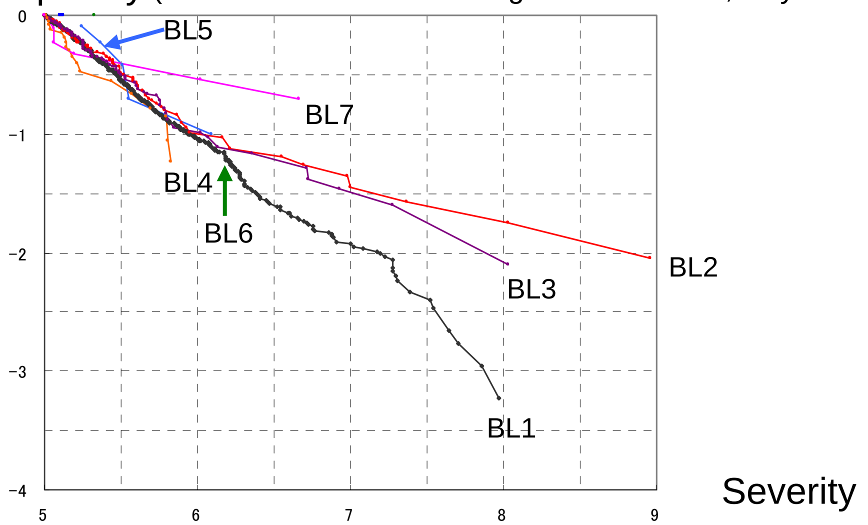
Graph 1

Frequency (number of losses greater than each threshold)



Graph 2

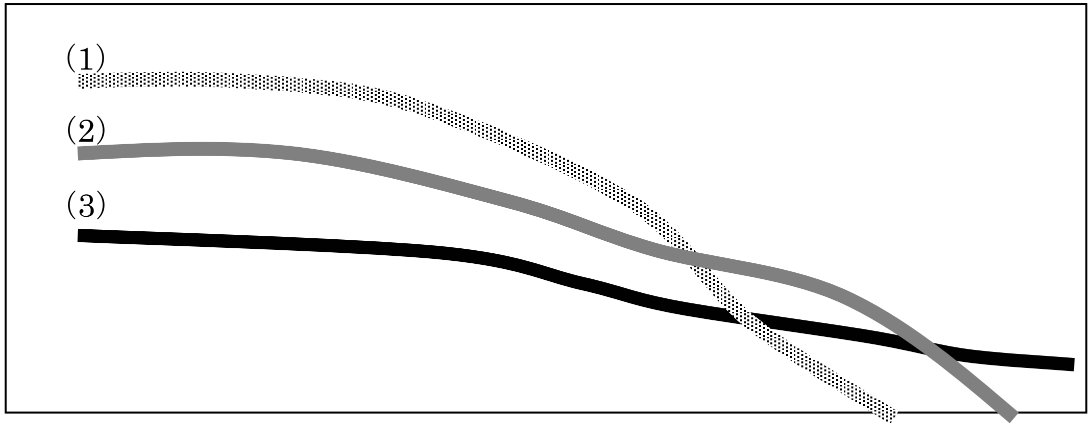
Frequency (fraction of losses exceeding each threshold, only for losses equal to or greater than 100,000 yen)



(3) Losses by type of operation

<1> Branches (domestic): branches, subbranches, …
<2> Others (domestic): headquarters, headquarter sections (dealings, M&A, …), centralized operations centers, …
<3> Overseas: subsidiaries, branches, …

- Observations:
  - ✓ <1> is the most frequent but appears to be the least risky.
  - ✓ <3> is the least frequent but appears to be the most risky.

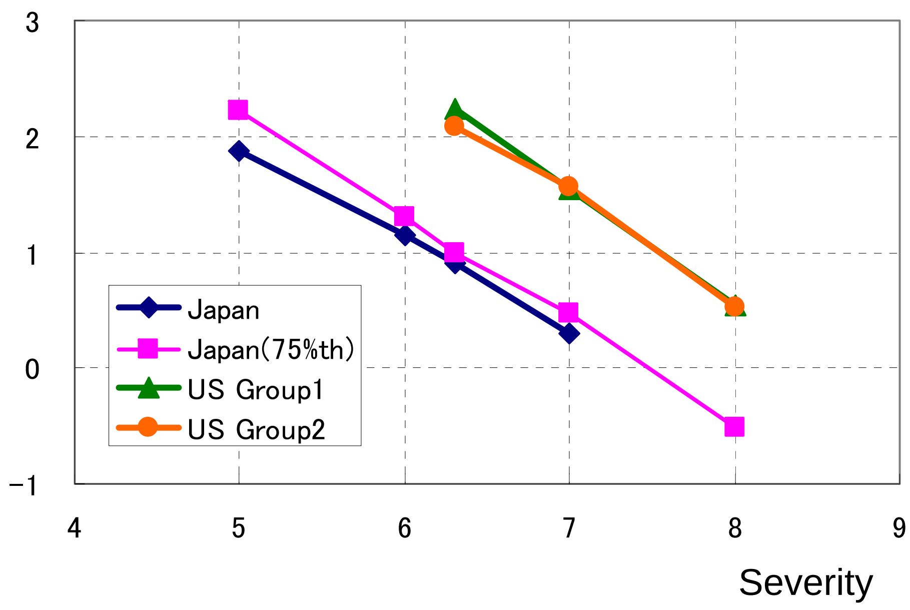


(4) Comparison between Japanese and U.S. banks' data

■ Comparison between the Japanese LDCE and the U.S. LDCE
($1 = 100 yen, upper stand is cross-firm median, lower stand is interquartile range)

	Losses ≥ ¥100K ($1K)	Losses ≥ ¥1M ($10K)	Losses ≥ ¥2M ($20K) Losses ≥ $20K	Losses ≥ ¥10M ($100K) Losses ≥ $100K	Losses ≥ ¥100M ($1M) Losses ≥ $1M
Japan LDCE 2007	76◆ (51–192■)	14◆ (8–20■)	8◆ (5–10■)	2◆ (1–3■)	0◆ (0–0.3■)
US LDCE 2005 Group 1 Institutions Reporting ≥ 1,000 Losses			176▲ (153–218)	35▲ (31–40)	3.5▲ (2.2–4.6)
US LDCE 2005 Group 2 Institutions Reporting ＜ 1,000 Losses			123● (91–210)	37● (10–44)	3.3● (0.0–3.8)

Frequency divided by total assets in 10trillion yen ($100billions) (See appendix for other scaling).



- ✓ Power law applies to both data sets.
- ✓ Curves from both LDCEs run parallel; at all points, the US LDCE data show a frequency about 15–20 times greater than that seen in the Japan LDCE.
- ✓ If the extrapolations are valid to the tail, the risk of a typical bank in the US LDCE is 15–20 times greater than that of a typical bank in the Japanese LDCE, when scaled to their total assets.

4. Conclusions

- Operational loss data from different institutions/countries share astonishingly similar characteristics.
  →This supports the importance of utilizing external loss data for individual banks.
  - ✓ While locations of "Double logarithmic graph" are different, in both countries operational risk losses follow the "power law".
  - ✓ The locations of the curves are very similar among Japanese banks.

- The "Double logarithmic graph" provides different kinds of information including data collection, risk profiles and approximate risk amounts.

- Additional information is required about tail losses, which drive the risk amount.
  - ✓ Further accumulation of loss data is needed in order to interpret the gentler slope at the tail of the "Double logarithmic graph" (whether it is because of the nature of the operational risk or it is the result of the paucity of the loss data).
  - ✓ For the purpose of risk management, in-depth analysis of individual loss cases may be more important.
    →Although this point is not dealt with in this presentation, it is more important to investigate tail losses (and small losses or near misses that could have turned into tail losses). Whereas Japanese banks have made great efforts to learn from internal loss data, they have not necessarily paid much attention to near misses and external losses.

Thank you.

Appendix

(1) Assumptions for graphical estimation of risk amount
(2) Comparison between the Japanese and the US LDCE (other scaling)
(3) References

(1) Assumptions for graphical estimation of risk amount
(a) Single loss approximation

Intersection of the extrapolation of the graph and the line parallel to the X-axis gives the approximation of the risk amount at the confidence level shown on the Y-axis.

(i) For example, the intersection of the graph and the line parallel to the X-axis (Y-value is –3) represents the largest loss in 1000 (10^3) years.

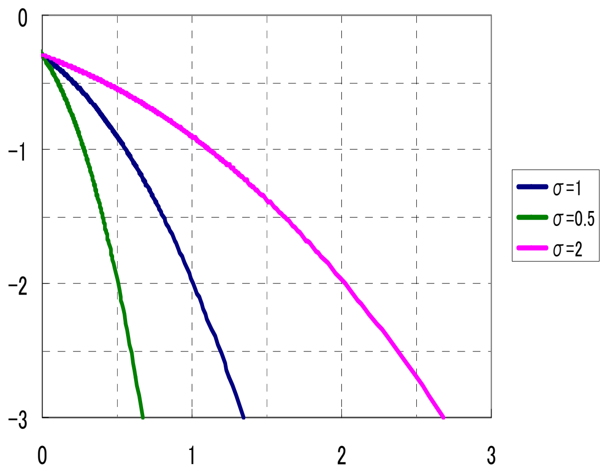
(ii) This is not the total amount of losses in the worst year (in this case it represents the 99.9% confidence level, i.e., the worst in 1000 years), as this is only the “single” largest loss.

(iii) However, when the loss data are heavy-tailed (which in practice they are), the single largest loss provides a good approximation of the total amount of losses in the worst year at high confidence levels (in this case, 99.9%).

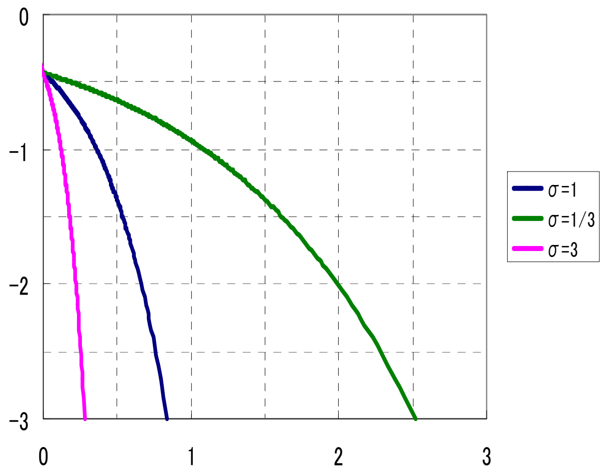
(b) Straight line extrapolation

Straight line extrapolation is valid as an approximation for many distributions, at least in the tail area.

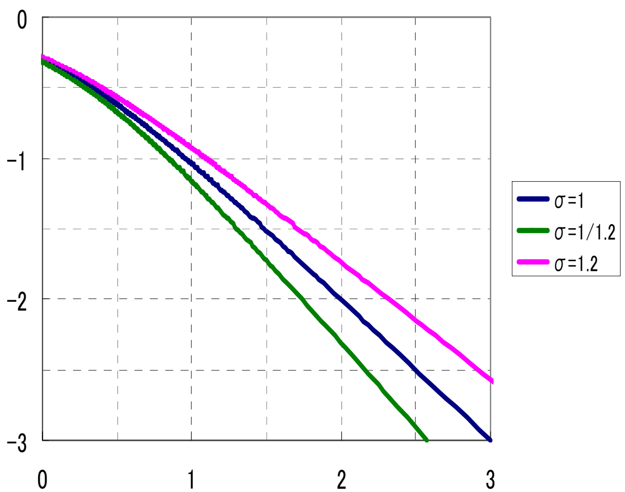
Log-Normal



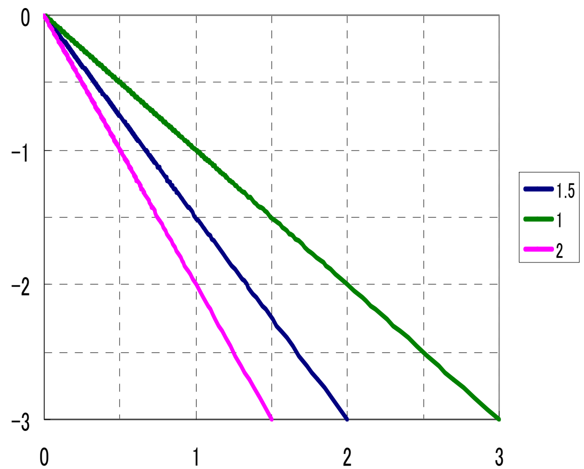
Weibull



GPD



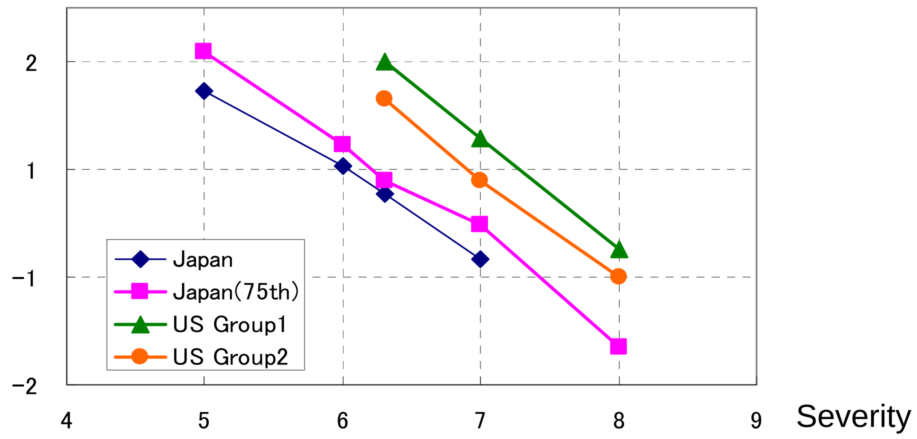
Power Law



(2) Comparison between Japanese and US LDCEs (other scaling)

(i) Loss frequency scaled to Tier1 capital

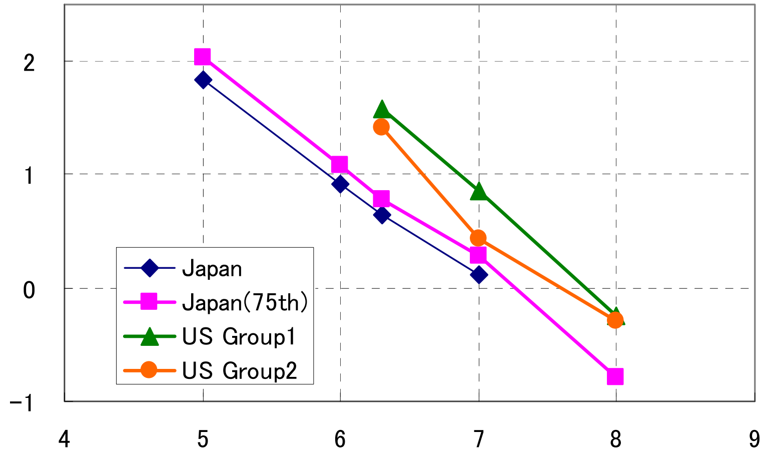
Frequency divided by Tier1 capital in 100 billions yen ($billions)



✓ The distance between the Japanese data and the US data narrows (to about 5–8 times).

(ii) Loss frequency scaled to gross income

Frequency divided by gross income in 100 billions yen ($billions)



✓ The distance between the Japanese data and the US data narrows further (to about 2–8 times).

(3) References

[1] Mori, A., T. Kimata and T. Nagafuji, "The Effect of the Choice of the Loss Severity Distribution and the Parameter Estimation Method on Operational Risk Measurement – Analysis Using Sample Data – ," Bank of Japan, 2007.

Appendix 1 of this paper gives a short explanatory note about "Single loss approximation" (Closed-form approximation) and numerical examples that examine the accuracy of the approximation of this method, using the loss data in financial institutions.

[2] Böcker, K. and C. Klüppelberg, "Operational VaR: A Closed-form Approximation," *Risk*, 18(12) pp. 90-93, 2005.

Includes short explanatory note about "Single loss approximation" (Closed-form approximation).

[3] Dekoker, R. and J. Newberry, "AMA Implementation at Citigroup Where We Are and Outstanding Questions," 2005.(Presentation material at "Operational risk … conference" at Boston federal reserve in 2005.

Includes analysis using "Double logarithmic plot" and numerical examples that illustrate accuracy of the approximation of this method.

[4] Dekoker, R., "Operational Risk Modelling: Where Do We Go From Here?" in *The Advanced Measurement Approach to Operational Risk*, pp. 37-57, 2006.

Includes detailed explanation of [3].

[5] de Fontnouvelle, P., J. Jordan and E. Rosengren, "Implications of Alternative Operational Risk Modeling Techniques," NBER Working Paper No. W11103, 2004.

Includes an analysis on US LDCE data using the "Double logarithmic plot"

(Note) [1], [3] and [5] can be obtained from the Internet.