Integrating Internal and External data

Sanpaolo IMI perspective

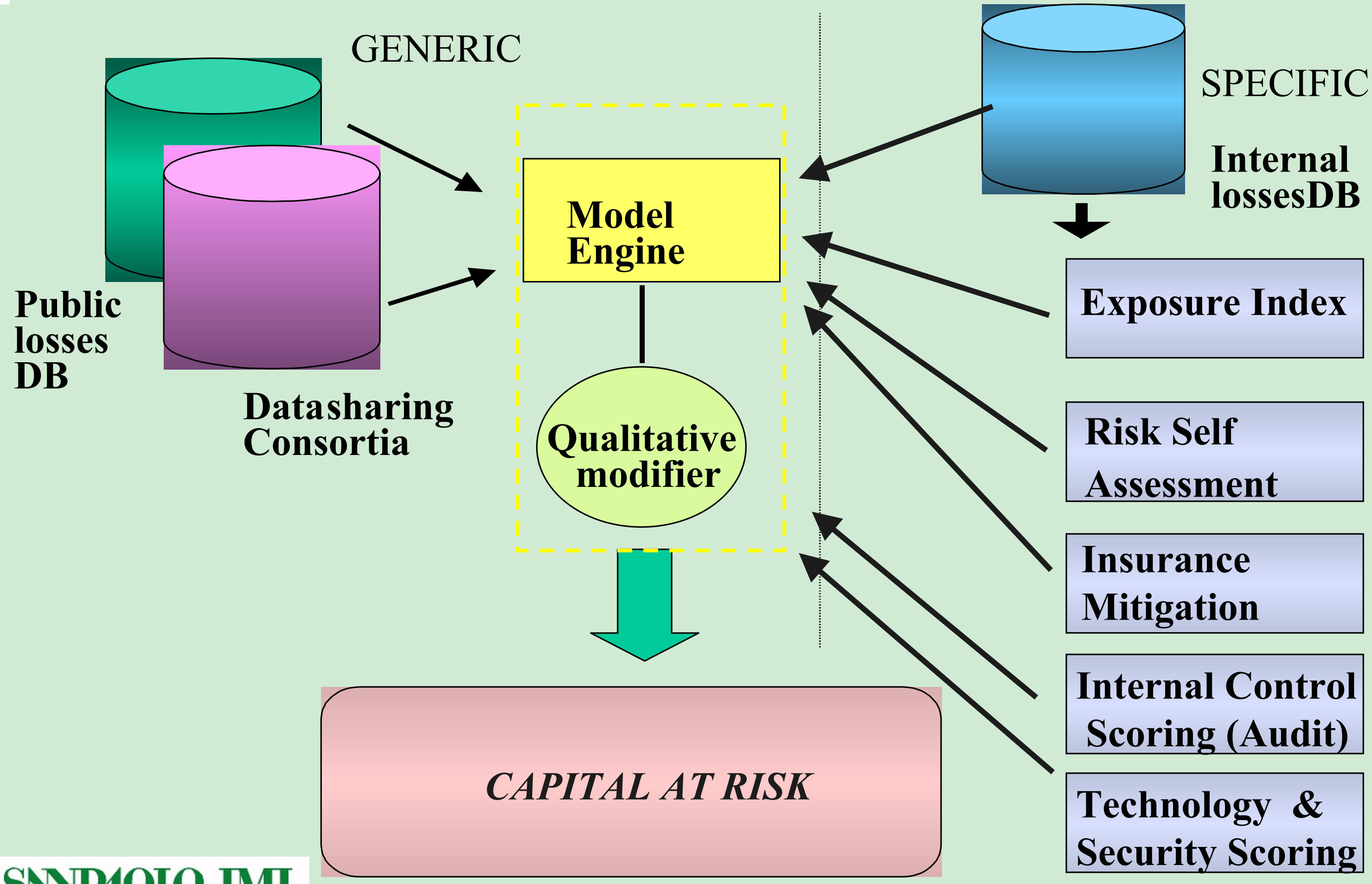
Giulio Mignola
New York, May 29-30, 2003

GRUPPO SANPAOLO IMI

Agenda

- Sanpaolo IMI approach for OpRisk
- What is external data?
- Difference in frequency and severity
- How about scaling?
- Integrating RSA and LDA results
- Conclusions

Sanpaolo IMI approach for OpRisk measurement



GENERIC

Public losses DB

Datasharing Consortia

Model Engine

Qualitative modifier

CAPITAL AT RISK

SPECIFIC

Internal lossesDB

Exposure Index

Risk Self Assessment

Insurance Mitigation

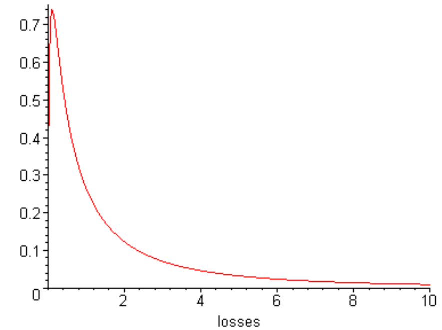
Internal Control Scoring (Audit)

Technology & Security Scoring

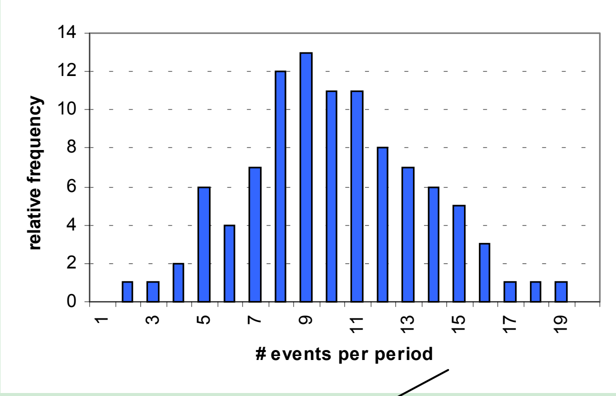
GRUPPO SANPAOLO IMI

Use of loss data: the LDA

Severity distribution

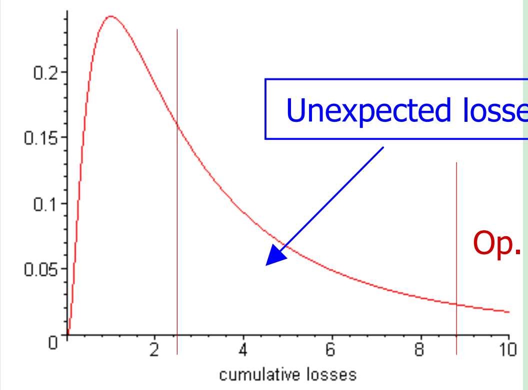


Frequency distribution



Aggregated via Monte Carlo Simulation

Total Loss Distribution



Unexpected losses

Op. VaR 99.9 percentile

Frequency analysis

- Study of how many events will happen in a given time period (e.g. 1 week)
- Typical probability distributions used:
  - Poisson
  - Negative Binomial

Severity analysis

- Study of the distribution of the impact of a **single** loss event independently from the period in which it happened
- Typical probability distributions used:
  - Weibull
  - Lognormal
  - Inverse Gaussian
  - Extreme Value Theory
- Estimation Tecniques:
  - Method of moments
  - Method of the maximum likelihood

Risk Self Assessment approach

- Use ⇒ “a priori” informations on the distribution
  - Frequency
  - Severity
- Ask questions to Business Managers about:
  - Expected number of events
  - Average Loss
  - Worst case scenario (in €’s)

$$\Downarrow$$

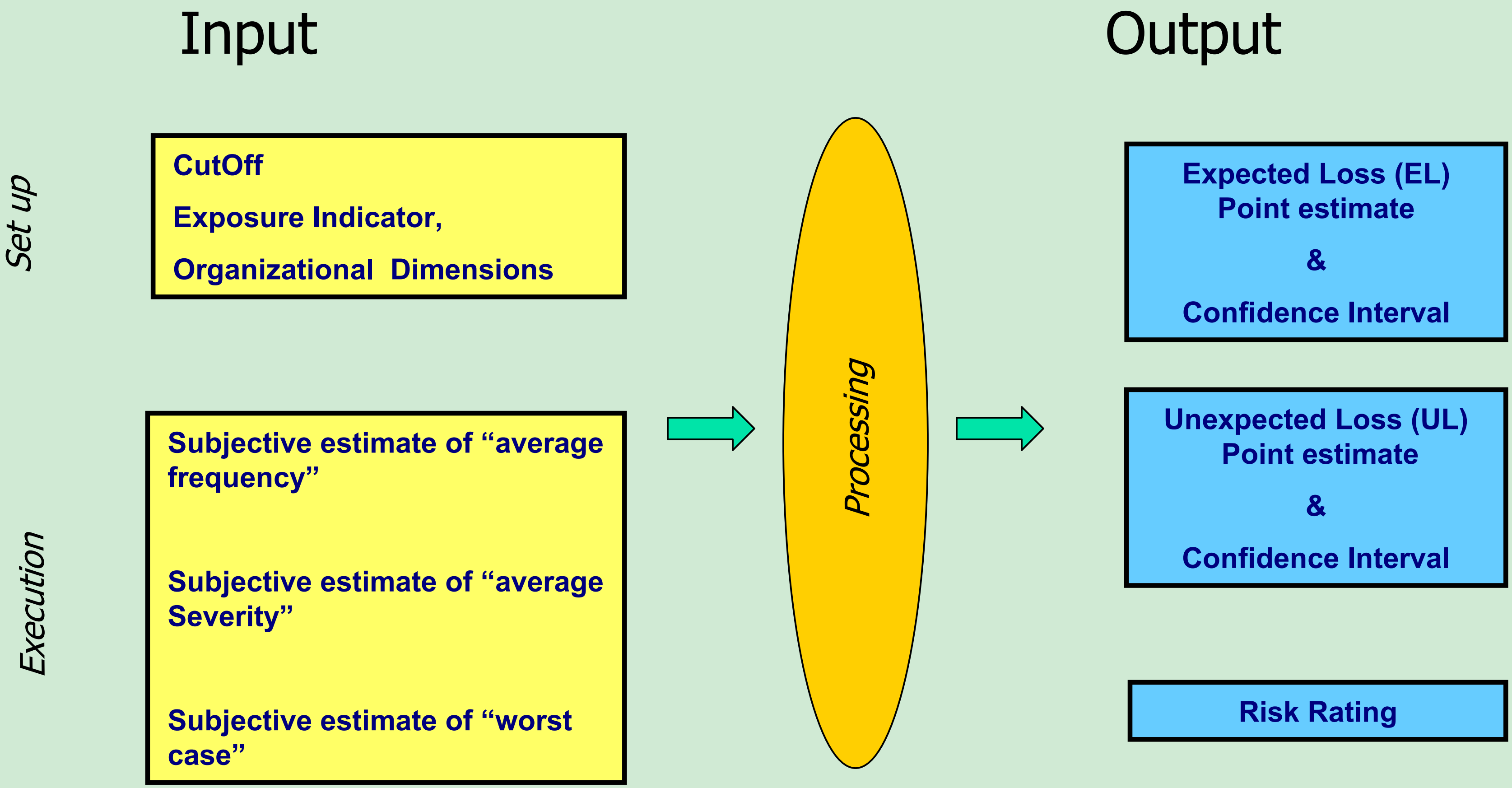
Deduct **C.A.R.**$^{\mathbf{RSA}}$

Risk Self Assessment: Objectives

- The methodology has been developed to:
  - Integrate scarce historical data and include a forward looking aspect in the measurement framework
  - Analyse risk factors and organizational dimensions in order to promptly address mitigation initiatives

- Solutions:
  - integrate historical loss data with subjective estimates consistent with statistical models
  - integrate the analysis with:
    - Key, Risk & Exposure Indicators
    - scenarios

RSA: input & output



Input

Output

Set up

CutOff

Exposure Indicator,

Organizational Dimensions

Execution

Subjective estimate of "average frequency"

Subjective estimate of "average Severity"

Subjective estimate of "worst case"

Processing

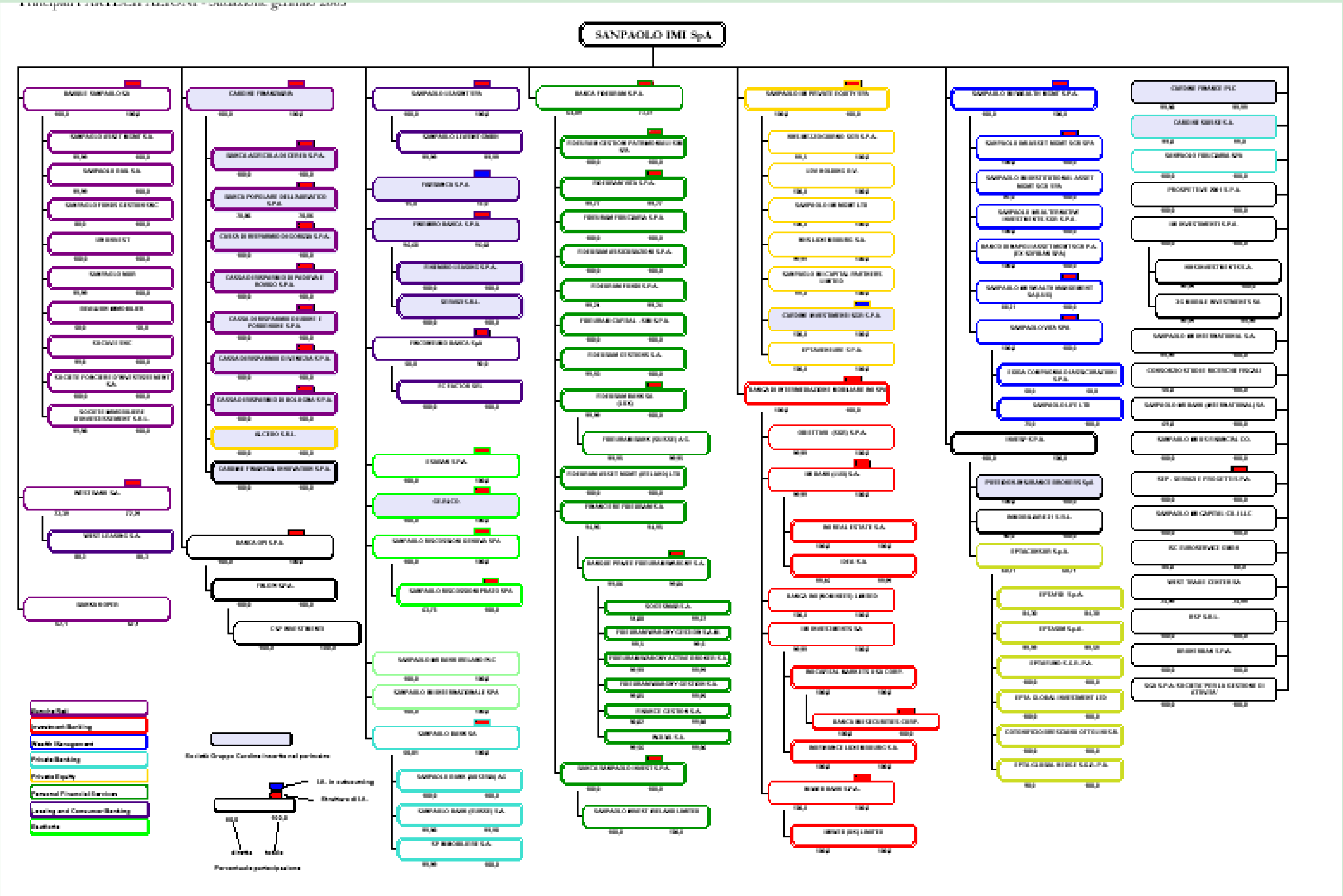
Expected Loss (EL)
Point estimate
&
Confidence Interval

Unexpected Loss (UL)
Point estimate
&
Confidence Interval

Risk Rating

How to deal with internal / external data mixing ?

A “typical” Group structure …..



… contains ….

- Several retail banks
- Some Asset Management firms
- An Investment bank
- Other support/product specific legal entities

All of which have

- Specific organizational structures
- Specific missions and responsibilities

  and
- Operate in a specific geographical/business environment

... so external data is ...

- Losses of bank n.1 w.r.t. the rest of the Group
- Losses of the Asset Management in France w.r.t. the Italian operations
- Losses of the recently acquired bank
- Etc. ......
- .... and of course loss data coming from (1) a consortium or (2) a public provider

… but also evolution (time) is important

- Losses of 2001 w.r.t. losses of 2002
- Losses of the retail brokerage operations “before” and “after” the introduction of the new STP system
- Etc……

How external data can be different from internal data from a “statistical” point of view

- Come from the same population, but are selected with different criteria (e.g. different threshold)
- Come from a different population, but with definite relations (i.e. scaling) with the internal data population (same distribution with different parameters)
- Are completely different (the loss generating mechanisms are different)

Data from the same population but with different threshold

- Dealing with point 1) is relatively easy, if the different threshold is known (and in the case of data pooling consortia it should be), the estimation procedure for the parameters can take it into account (e.g. via the maximum likelihood methodology)

Maximum Likelihood Estimation with threshold

- In general the maximum likelihood function to be maximized w.r.t. θ is:

M.L. Function = $\prod_i f(x_i;\theta)$

- With a threshold $x_{(th)}$ it becomes:

$$\text{M.L. Function}_{(th)} = \frac{\prod_i f(x_i;\theta)}{1 - F(x_{(th)};\theta)}$$

2) Data with the same distribution but different parameters

- In this case a so called "scaling" relation has to be implemented.
  - **N.B.** Scaling formulas must be derived for each bank and nothing ensures that scaling formula can be imported from one bank to another
- The dynamic of the loss generating process has to be considered carefully in order to properly treat the different components
  - Frequency scaling
  - Severity scaling

Are then frequency and severity different? (1)

- Frequency: how many loss events take place in a given period of time
  - Factors relevant to Frequency:
    - Size of the institution/B.U.
    - Line controls (first level control system)
    - "Environment" (internal/business/social)

Are then frequency and severity different? (2)

- Severity: once it happens, how large is a loss?
  - Factors relevant to Severity:
    - Size of the single “transaction”
    - Secondary controls (second and third level)
    - Insurance coverage
    - “Environment” (internal/business/social)

Analysis (1)

- But all of the previous factors can act differently by event type

Example: legal liabilities arising from contractual problems in mutual funds selling

  - Frequency → the number of different products that the firm sells
  - Severity → the number of customers and the balance sheet of the firm (mass litigation)

Analysis (2)

- Another example: robberies in a retail network
  - Frequency → the number of branches
  - Severity → the cash amount inside the branch (this is a typical geographical issue – the more electronic means of payment used, the less cash inside the branch)

A first example of “scaling”…

- Event type: Robberies
- Number of robberies proportional to the number of branches
- Frequency: Poisson distribution

$$\lambda_B = \lambda_S * N_B / N_S$$

where λ_S is the average number of robberies per branch per year in the “system” and N_S is the total number of branches in the “system” (subscript B refers to the single bank)

…Another proposal[1] (scaling of severity)

$$Scaled\ Loss = L_{DB}\left(\frac{R_{Bank}}{R_{DB}}\right)^{\alpha}$$

L_{DB} = Actual Loss experienced by bank

R_{Bank} = Current Revenue of bank

R_{DB} = Previous Revenue of bank

α = Scaling coefficient determined by regression analysis

in general:

$$E(x) = f(Size, G.I., Total\ Asset)$$

where $x = x_0 \cdot (Size\ Ratio)^{\alpha}$ and $\alpha \approx 0,23$

[1] Shih, J., A. Samad-Khan and P. Medapa, "Is the size of an Operational Loss Related to Firm Size", Operational Risk (January 2000) and A. Samad-Khan "An integrated Framework for Measuring and Managing Operational Risk" , OpRisk 2003

3) Data from completely different populations

- In this case, the hypothesis is that the loss generation mechanism of the external data source is different from internal one.

  We can take two positions:

  - The external mechanism is so “rare” that in the internal data we do not find it (but it is there)
  - The internal environment is free from that type of losses (e.g. because of business mix and operating processes)

What happens in 3a. Case?

Problem:

- Frequency is very low
- This implies that the only significant severity comes from external data

e.g. ***a mass litigation due to a supposed mutual fund misselling in case of market downturn:***

- you might never have experienced one
- but, if you are into fund selling business, certainly you are exposed to

A note about the importance of tail events

- LDC 2002 shows that:
  - 0.2% (63 out 37.000) of of all reported losses accounts for more than 50% of the total loss amount
  - Losses above 1.000.000 € account for more than 70% of the total loss amount and represent 1.3% of the total number of losses (482 out of 37.000)
  - Evidence from data should be taken with some caution because of the <u>distortion effect induced by the 9/11 event</u>

Some considerations

- In a real world situation, cases 1) to 3) are mixed and case 3b) is very difficult to justify
- In the case of a consortium and “a fortiori” in the case of data coming from different parts of the organisation, some scaling may be tested against data (e.g. the number of robberies vs the number of branches, etc.)
- For tail events, mixing with case 3a) should be considered via the application of credibility theory

Credibility (in a more general sense)...

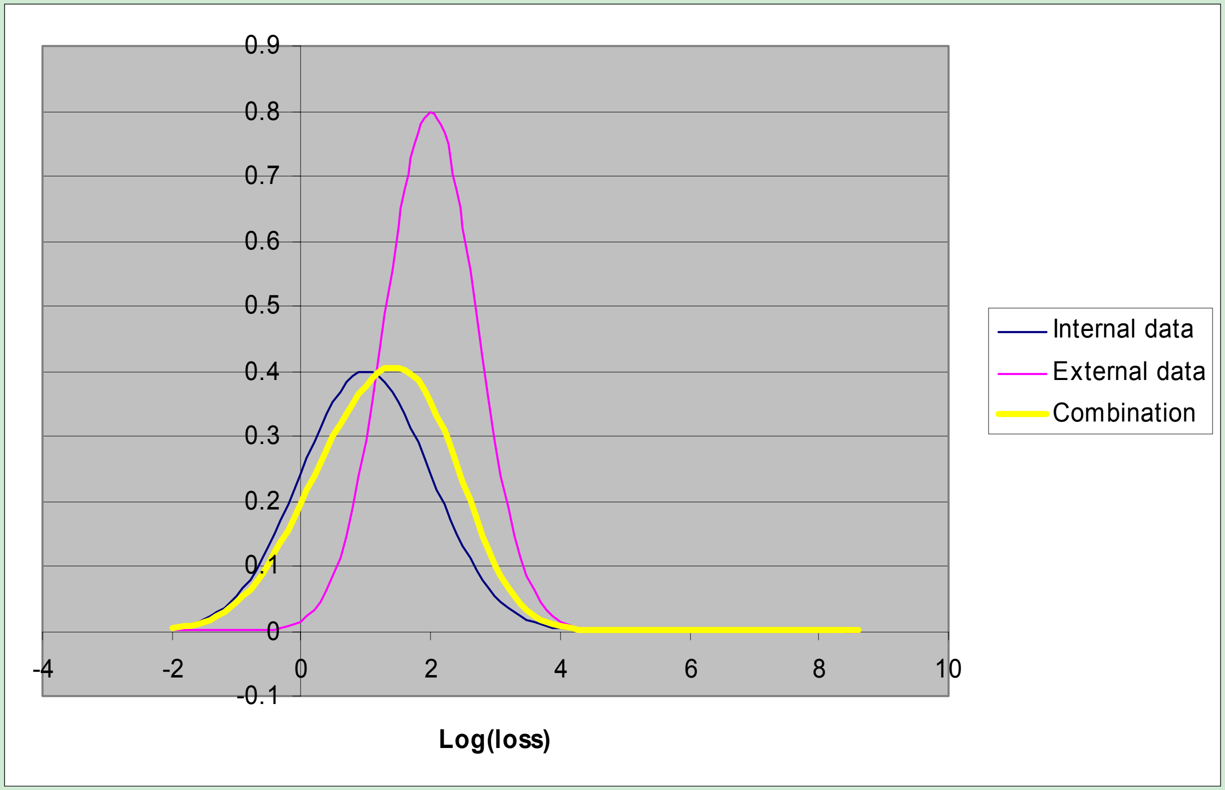
$$A_{eff} = W \times A_{\text{int}} + (1 - W) \times A_{ext}$$

where

- A is a general quantity related to the measurement process (e.g. the loss distribution) and
- W is a credibility weight function that takes value in [0,1] and depends on:
  - Size of the samples (internal and external),
  - judgment (prior knowledge) on the dynamic of the loss generation process,
  - business environment,
  - internal control system
  - etc.

Illustrative example for loss data

Weight on internal data W = 0.8



Same technique to integrate RSA and LDA results

- RSA $\Rightarrow$ Capital [C.A.R. ^{RSA}]
- LDA $\Rightarrow$ Capital [C.A.R. ^{LDA}]

$$\Downarrow$$

$$C.A.R. = Z \times C.A.R.^{RSA} + (1 - Z) \times C.A.R.^{LDA}$$

where Z is a function of based on prior knowledge on:

- size
- time stability
- “environment”
- and hence of **judgement** of *Operational Risk Managers*

GRUPPO SANPAOLO IMI

Conclusions (1)

- External data is unavoidable
- Analysis of relevance is mandatory because external data can be, for the reasons explained before, extremely different from internal one
- One (*or more ?*) kind of “statistical” integration of internal and external data can successfully be implemented
- The data collected in the “Loss Data Collection” exercise is extremely valuable and it is highly important that such exercises will continue in the future in order to analyse Op Risk aspects and hence allow all banks to capitalize on data they will never (or very difficult) get otherwise

Conclusions (2)

- A risk self assessment is the element that complete the measurement framework. Adding a measure of risk where data is lacking and/or not representative of the bank/unit;
- The integration process between RSA and LDA could be done by using exactly the same methodology adopted for the integration between internal and external data