MODELLING OPERATIONAL RISK FROM LOSS DATA

THE UNRESOLVED CHALLENGE

Giulio Mignola
Intesa Sanpaolo

Industrial-Academic Forum on Operational Risk - Toronto, March 26th ,27th

Disclaimer

Opinions and views expressed here are of the speaker exclusively and do not necessarily represent practice or policies of Intesa Sanpaolo Spa

Summary

- The regulatory challenge: what are banks asked to measure?
- Idiosyncrasies of operational risk loss distributions
- Pitfalls in common approaches to estimating regulatory capital
  - Extrapolating beyond the range of the data
  - Under-determination of models by data
  - Level playing field ?
- The need for quality data
- A proposal for modesty

Understanding some challenges posed by Basel II

The BIS tower in Basel

- In the Basel Accord on Capital Adequacy Operational Risk is identified as a distinct source of banking risk and a specific capital requirement is introduced
- In addition to basic and standard measurement methodologies, stringent requirements for the internal models are defined.
- The crucial quantitative requirement characterizing the AMA is to produce a <u>risk measure which is "compatible" with the 99.9% of the annual loss distribution</u> (i.e. the distribution of the total loss of one year).
- Another stringent requirement is the collection of loss data (both internal and system) that covers at least 5 years history

Regulatory treatment of different risks

- Credit risk (for exposures in corporates)

K =1,06 ***LGD** *{ N [(1 – R)^–0,5 * G (**PD**) + (R / (1 – R))^0,5 * G(0,999)]– **PD**} * [1 + (M – 2,5) * b]/ (1 – 1,5 * b)

PD and LGD determined using internal models, formula to obtain K is fixed by the Regulation (PD*LGD is substantially the expected loss)

- Market Risk (for linear instruments)

$K = \mathbf{VaR_d^{99}} * (10)^{0.5} * R$

Daily VaR at 99 percentile, scaled on 10 days horizon (Sqrt(10) factor) times a "regulatory factor" (between 3 and 4 depending on performance of backtesting, quality of processes etc.)

- Operational Risk

$K = \mathbf{F_{1y}^{-1}}(0.999)$

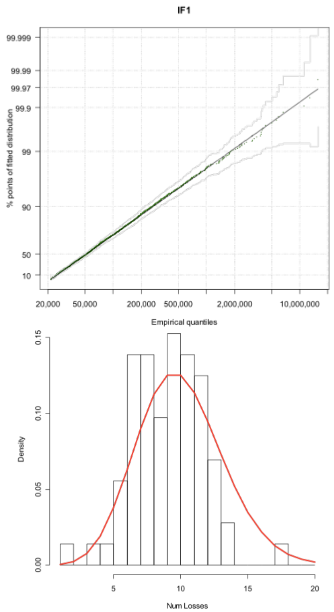
Risk measure which is "compatible" with the 99.9% of the annual loss distribution (i.e. the distribution of the total loss of one year). The distribution itself is determined by the bank !

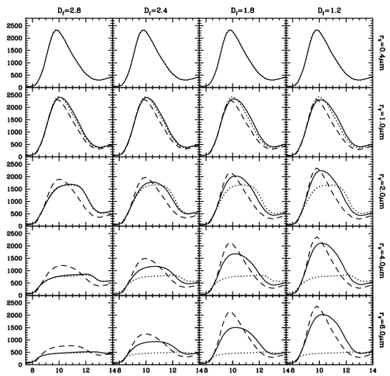
Bank internal estimate

The Loss Distribution Approach (LDA)

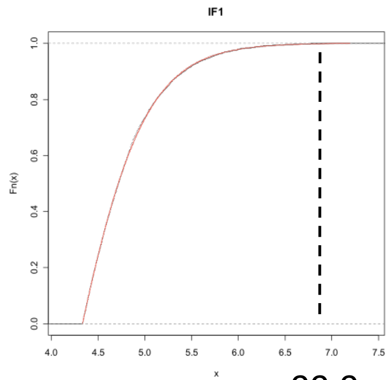
LDA is an actuarial, bottom-up approach to computing regulatory capital

1) Estimate individual loss frequency and severity distributions for each unit of measure
2) Combine frequencies and severities to determine annual loss distributions
3) Estimate correlations among dependencies among annual UOM losses
4) Compute the total annual loss distribution and estimate high percentiles





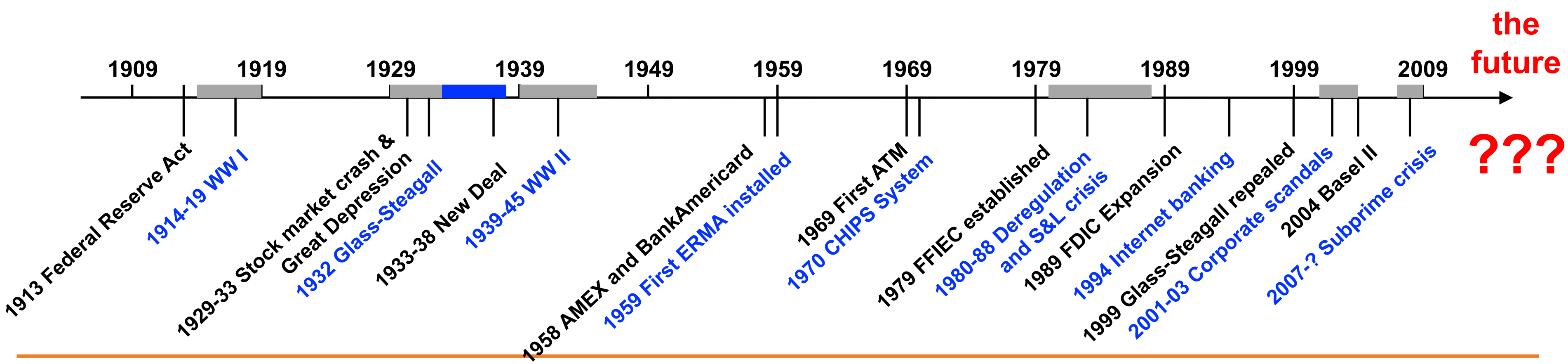
--	0.02	0.05	0.06	0.02	0.00	0.05	0.20
0.02	--	0.03	0.02	0.08	0.13	0.10	0.01
0.05	0.03	--	0.07	0.07	0.04	0.04	-0.13
0.06	0.02	0.07	--	0.07	0.12	0.05	0.02
0.02	0.08	0.07	0.07	--	-0.01	0.00	0.24
0.00	0.13	0.04	0.12	-0.01	--	0.10	0.08
0.05	0.10	0.04	0.05	0.00	0.10	--	0.12
0.20	0.01	-0.13	0.02	0.24	0.08	0.12	--



- 99.9 %

The 99.9th percentile in perspective

- The 99.9th percentile represents a loss that would occur with an average frequency of **once per 1,000 years**
  - Assuming the distribution of yearly total operational losses for a bank remains constant
- No bank can directly estimate this quantity based on internal data alone
  - The AMA standard can only be used for regulatory purposes if the bank has collected a **minimum of 5 years** of internal loss event data
  - The oldest bank in the world is Monte dei Paschi di Siena, founded in 1472 – 537 years old
- Consider what has happened in the past 100 years alone in the US:

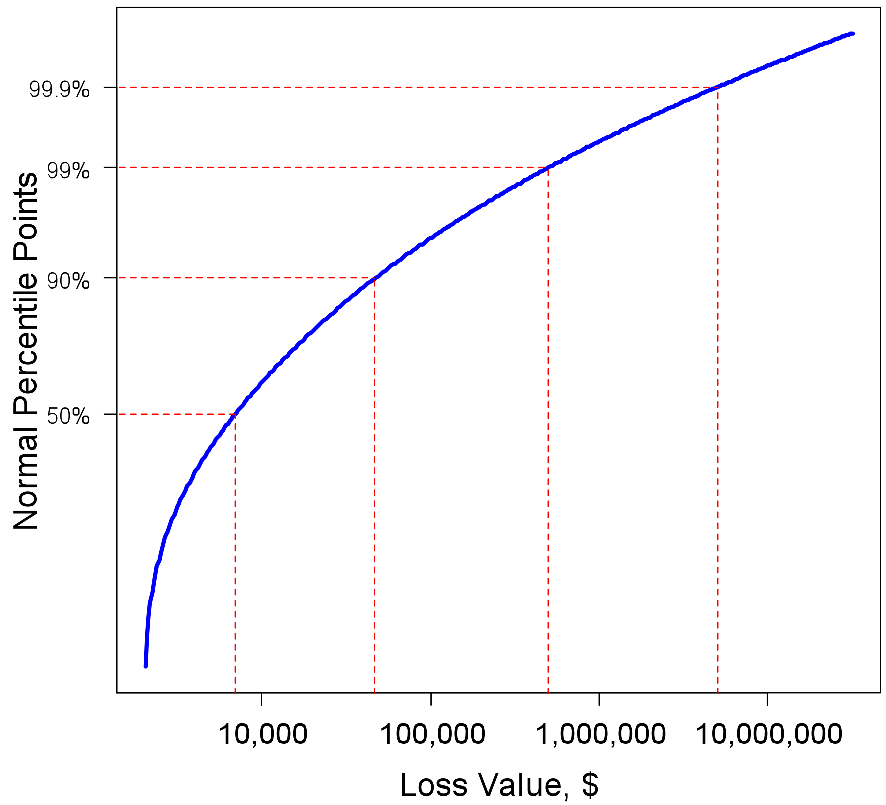


“Heavy tails” and their consequences

- Operational loss severity distributions typically show an “heavy-tailed” behaviour
  - “Typical” extreme losses can be several orders of magnitude larger than the median loss
  - N.B. we use “heavy-tailed” to refer to the distribution of *observed* losses, not the complete loss distribution

- Heavy-tailed losses tend to dominate other losses
  - *Instability of estimates:* Occurrence of an extreme loss can cause estimates of the mean or variance of losses to change dramatically
  - *Dominance of sums*: The sum of a set of losses is typically of the same order of magnitude as the largest single loss
  - *Dominance of mixtures:* If losses from a heavy-tailed category are mixed with lighter-tailed losses, the mixture will also be heavy-tailed

Extreme Values of a Heavy-Tailed Distribution



How much data do we need?

- Let's examine a typical severity distribution fitted to **individual** losses in Internal Fraud / Retail Banking for Western European banks
  - Suppose the losses are generated according to a Lognormal distribution with $\mu=10$, $\sigma=2$
  - The 99.9^{th} percentile of this distribution is **10.6 million Euro**

Number of data points	Width of a 95% confidence interval for the 99.9^{th} %ile
1,000	**20.0 M**
10,000	**7.7 M**
100,000	**2.5 M**
1,000,000	**0.8 M**

Estimating total losses for a unit of measure

- We often have lots of data on individual losses, but relatively few years' worth of data on total losses
  - Can we use estimates of single-loss distributions to derive distributions of total losses?
    - ► Yes, although complex calculations are typically required

- For heavy-tailed distributions, the *single-loss approximation (SLA)* provides an easy heuristic to deriving the distribution of total losses
  - Suppose a bank experiences 10 losses per year on average in one UOM
  - The 99.9th percentile of the sum of 10 losses is approximately equal to the 99.99th percentile of the single-loss distribution

- While convenient, there are some unpleasant implications of the SLA
  - We need ~ 10x as much single loss data to get a comparably accurate capital estimate
  - If we use single-loss distributions as the basis of a capital calculation, we effectively are using values extrapolated in the tail well beyond the 99.9% level

The single loss "approximation"

Following known results recently highlighted in *C. Klüppelberg and K Böcker, "Operational VaR: a Close-Form Approximation)*

For a large class of severity distribution (sub-exponential) it can be proven that

$$\lim_{x \to \infty} \frac{\overline{F^{*n}(x)}}{\overline{F(x)}} = n$$

And hence for a high percentile of the aggregated distribution the following approximation holds

$$1 - F_S(x) \approx \mathrm{E}[n](1 - F(x))$$

$$F_S^{-1}(p) \approx F^{-1}\left(1 - \frac{1-p}{E[n]}\right)$$

For example for an average frequency of 200 losses per year we have

$$VaR = F_S^{-1}(0.999) \approx F^{-1}\left(1 - \frac{1-0.999}{200}\right) = F^{-1}(0.999995)$$

the approx. VaR is the 99.9995 percentile of the severity (on average the largest event on a sample of 200,000)

How much extrapolation is required?

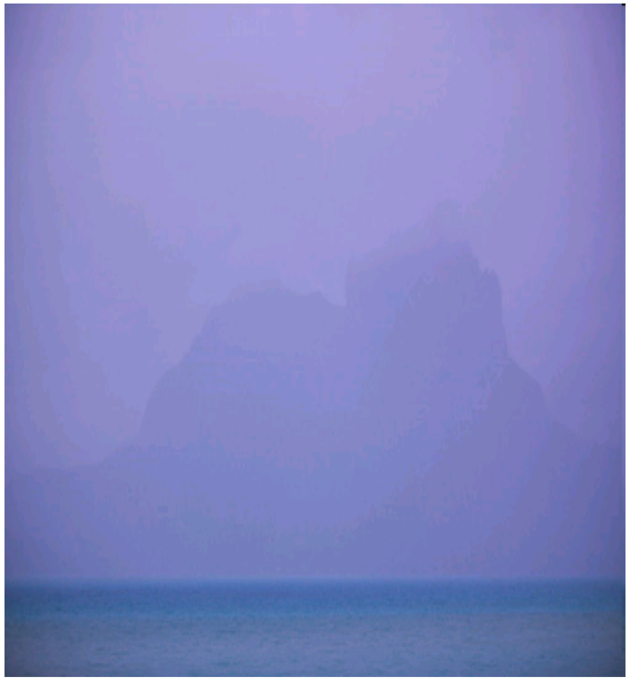
- The 99.9% annual loss confidence level with a 5-year loss history in the database requires a 1000/5 = 200-fold extrapolation beyond the observed data
  - External data from the industry could help a bit, but at least a data pool of 200 banks for 5 years is needed. Homogeneity of the various contributors to the pool remains questionable.
- To reach the 99.9 percentile of the annual aggregate loss we must explore regions of the loss severity which are far beyond the observed data points
  - Example: using the single loss approximation, if we want the 99.9 percentile of an aggregate loss obtained with a frequency of 100 losses per year, we need to estimate up to the **99.999** percent level, or the **1-per-100,000 loss event**
  - Recognizing that in a typical “risk class” we have at most a few thousand data points, it is almost sure that we need to guess the behaviour of losses in the tail

What can we say about losses beyond the range of data?

If only past losses are used as a guide, then we must find a way to extrapolate to the area of interest from observed losses

How legitimate is it to do this?

- We must assume that there are *no structural differences* in the way that losses are generated at high levels
  - This may not be the case if losses are generated in many ways within a unit of measure
  - **There may be hard limits to how much can be lost in a given incident**
  - Mitigation, recovery, and control mechanisms may function differently for large losses
- Even if this is true, the extrapolation error increases substantially beyond the observed data
- Additional justification may come from experimental evidence, but such evidence will be slow in coming



What lurks beyond the horizon?

“Hic sunt dracones”: Here there be dragons

What hope does Extreme Value Theory have to offer?

- Extreme Value Theory (EVT) is a very well-developed area of probability that concerns the distribution of large losses
  - Similar to what the Central Limit Theorem offers for computing averages, EVT offers us for computing distributions of maxima and extreme quantiles
  - EVT has worked well in hydrology, e.g., in determining how high to build dams

- Under some regularity conditions, EVT shows that the distribution of losses exceeding a high enough threshold value is a Generalized Pareto Distribution (GPD)
  - If the assumptions are true, then we have a valid means of extrapolating beyond our data!
  - The GPD takes a variety of tail shapes determined by a shape parameter ξ, to be estimated

- However, EVT is mathematics, not statistics. Statistically, we must ask:
  - How do we know the regularity assumptions are satisfied in the context of OpRisk data?
  - How can we infer from the data that we have selected a “high enough” threshold?

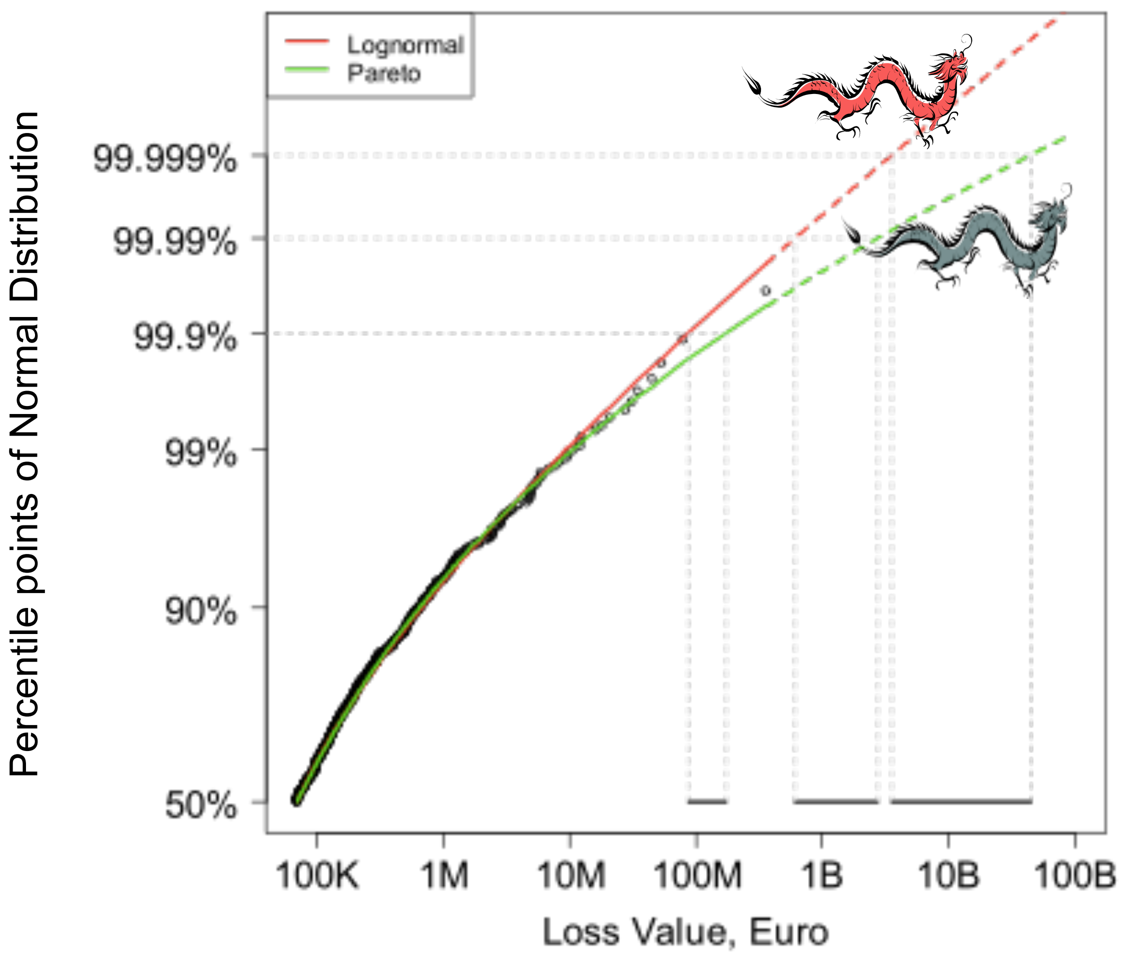
The EVT leap of faith

- There is no practical way of statistically determining whether either assumption has been met
  - At bottom, we must take it on faith that any given extrapolation is valid
  - The validity of any extrapolation must ultimately be empirically grounded

- Some distributions that meet the EVT criteria and fit observed OpRisk data very well require very high thresholds to get accurate EVT-based estimates
  - Mignola and Ugoccioni (2006) observed this for the lognormal distribution
  - Degen, Embrechts and Lambrigger (2007) have shown this is true for the g-and-h distribution (indeed Degen and Embrechts (2009) discuss alternative definitions for convergence at high quantiles and an improved P.O.T.-type formula)
  - Using GPD estimation on the observed data leads to extremely high capital estimates
  - This implies, to be on safe grounds, that we would need a lot more data before EVT would give us reliable estimates

- The GPD might be a good model for extreme loss events in some cases, but only if it has been empirically established, not because of asymptotic theorems

Choose your own adventure – two possible endings



Dominance of high-severity, low-frequency losses

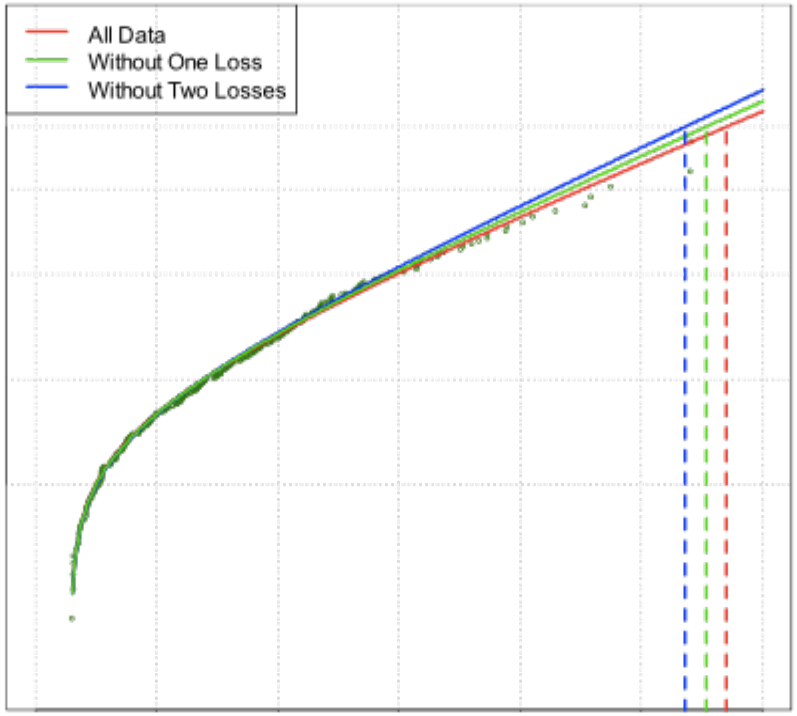
- The estimated value and uncertainty around the capital estimates of the LDA approach can be determined primarily by a few units of measure
  - High severity losses have higher capital values and wider margins of error
  - Low-frequency losses also have wide margins of error for the capital value

- A simple simulation experiment shows the magnitude of this effect
  - We simulated annual losses from 10 business lines, each having representative high-frequency, low-severity losses (1000 losses per business line), estimated the loss distribution for each BL, and then estimated total capital
    - The standard deviation of capital estimates is around 10% of the mean capital value
    - Each business line contributes 10% to the total error
  - We then replaced one of the ten loss distributions with a representative high-severity, low-frequency loss distribution (100 losses)
    - The standard deviation of capital estimates is now around 75% of the mean
    - The single heavy-tailed business line contributes almost 100% of this error

Sensitivity to the largest data values

- The occurrence of a single extreme loss can cause wild fluctuations in the estimated capital for a unit of measure
- Example: Corporate Finance / Clients, Products & Business Practices
- 99.9th percentile of a fitted lognormal distribution to the individual loss severities (using maximum likelihood estimates) :
  - All data: € 5,030 MM
  - Without maximum loss: € 3,412 MM
  - Without two largest losses: € 2,295 MM

Lognormal fits to severity data



INTESA SANPAOLO

The impact on level playing field

- Two main areas of concern
  - Fundamental issues
    - Are models able to distinguish two banks with different risk profiles ?
    - Or conversely, and far more importantly, are models capable of correctly representing two banks with the same risk profile by producing the same result ?
  - Regulatory issues
    - Does the validation process pose the same challenges from one jurisdiction to the next ?
    - Are the different blends of components (i.e. scenarios vs. loss data) allowed in different ways from one supervisor to the next really capable of representing the financial firms risk profiles' in the same way ?
    - Use of correlations?
- Of these two areas of concern, the second one, although being of extreme importance, is not connected with the properties of the Oprisk data and is of no interest here.

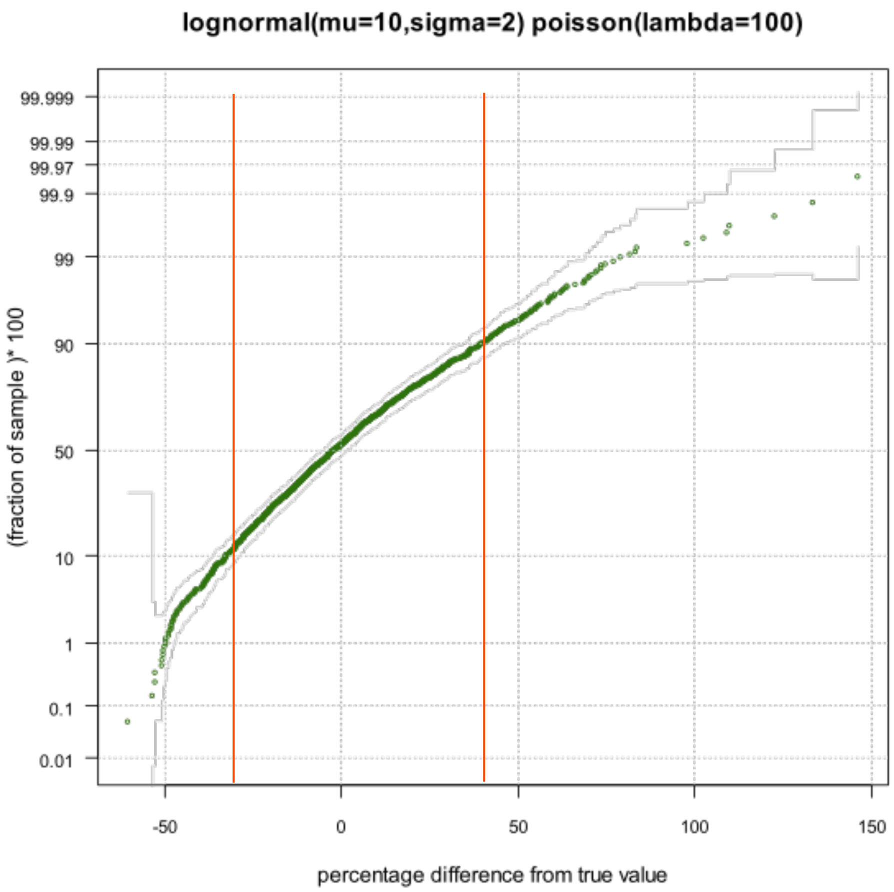
Are we really measuring what we're intended to ?

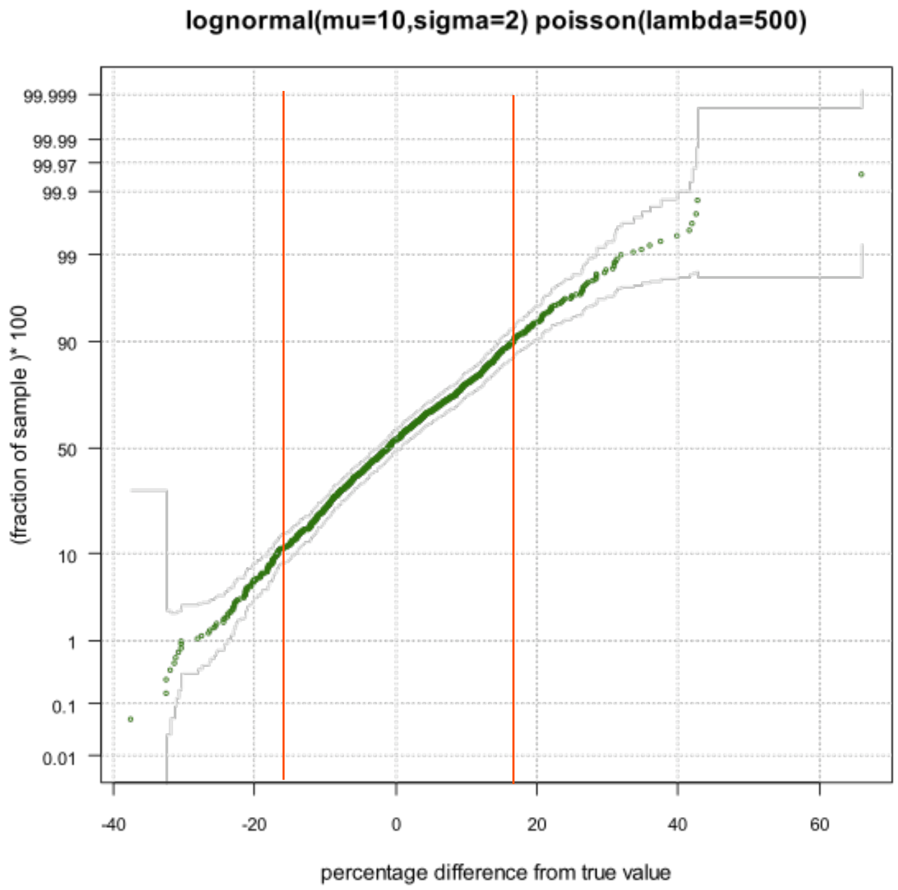
- Lets consider a typical unrealistic "gedanken experiment"
  - Define the risk profile of a given bank (mono business, mono risk) by defining the set of parameters of its underlying loss distributions (one for frequency – λ and a set for the severity distribution – $\boldsymbol{\theta}$).

    In this case the risk measure is easily defined as the 99.9 percentile of the aggregated annual distribution;
  - Consider a world in which only banks exist (say 1000 of them); identical in terms of their operational risk profiles (same λ and $\boldsymbol{\theta}$).

    In principle one should expect that each bank's internal model would produce nearly the same result in terms of level of risk .
  - Simulate the loss database of each bank (say 5 years of data, on average $5 \times \lambda$ data points per bank)
  - Apply the same model to estimate the risk profile from data and compare the results

Results – a very uncertain world

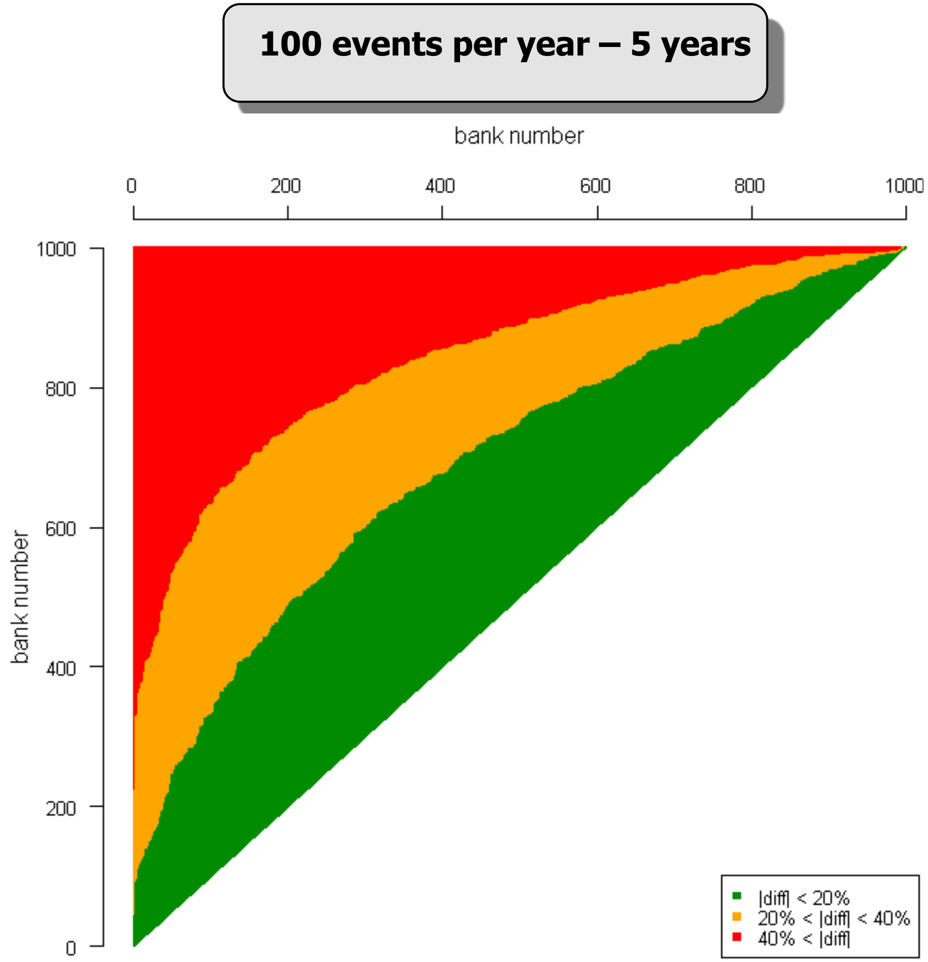


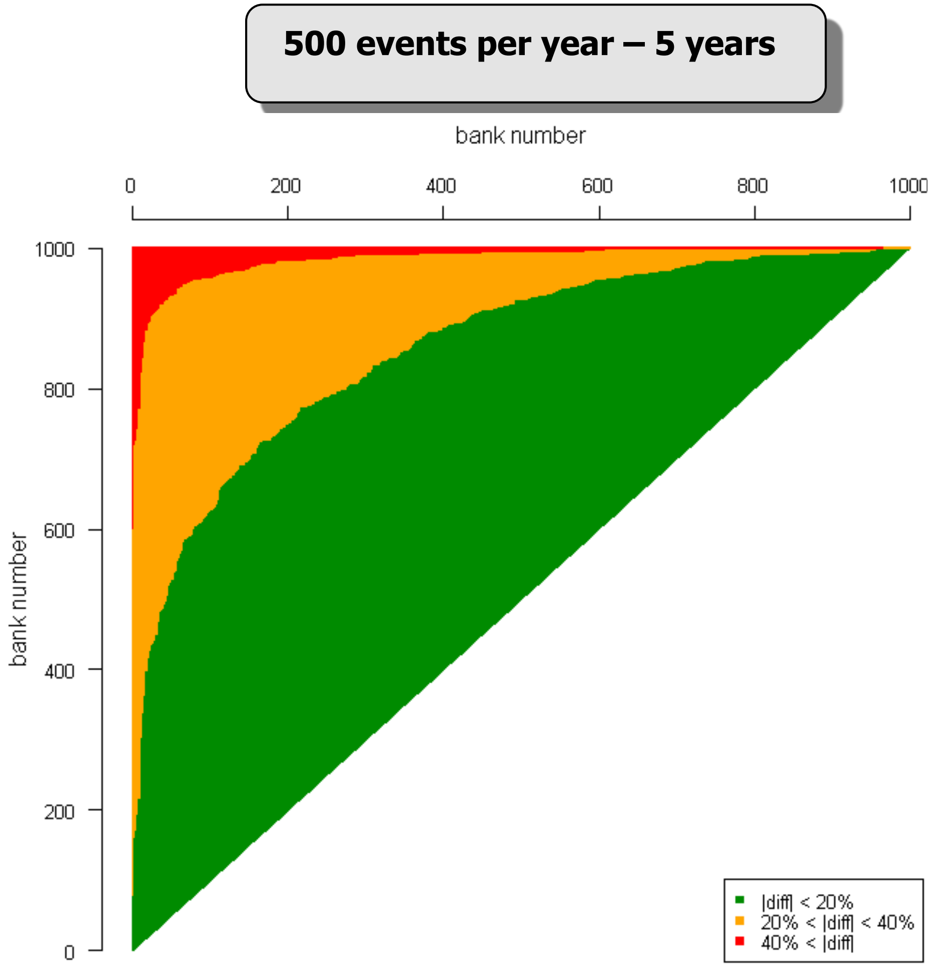


20% of the cases have differences greater than 35% from the true value (10% +35% and above, 10% -35% and below)

20% of the cases have differences greater than 18% from the true value (10% +18% and above, 10% -18% and below)

Results – a very unlevel playing field

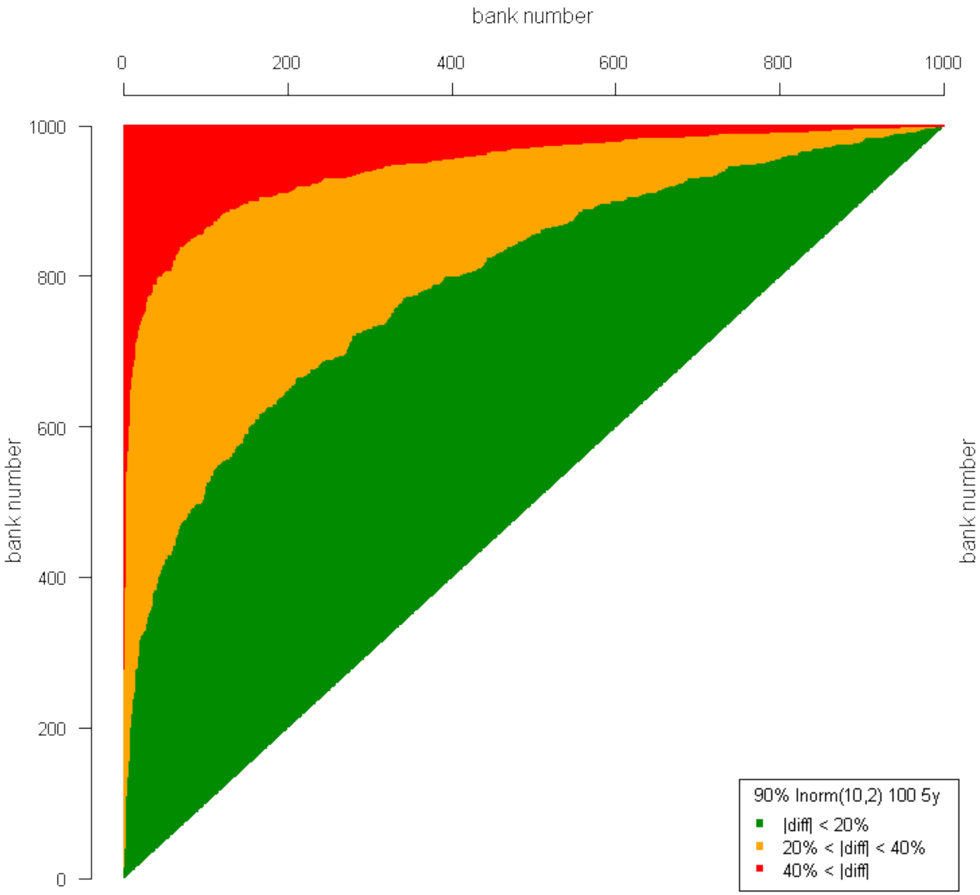




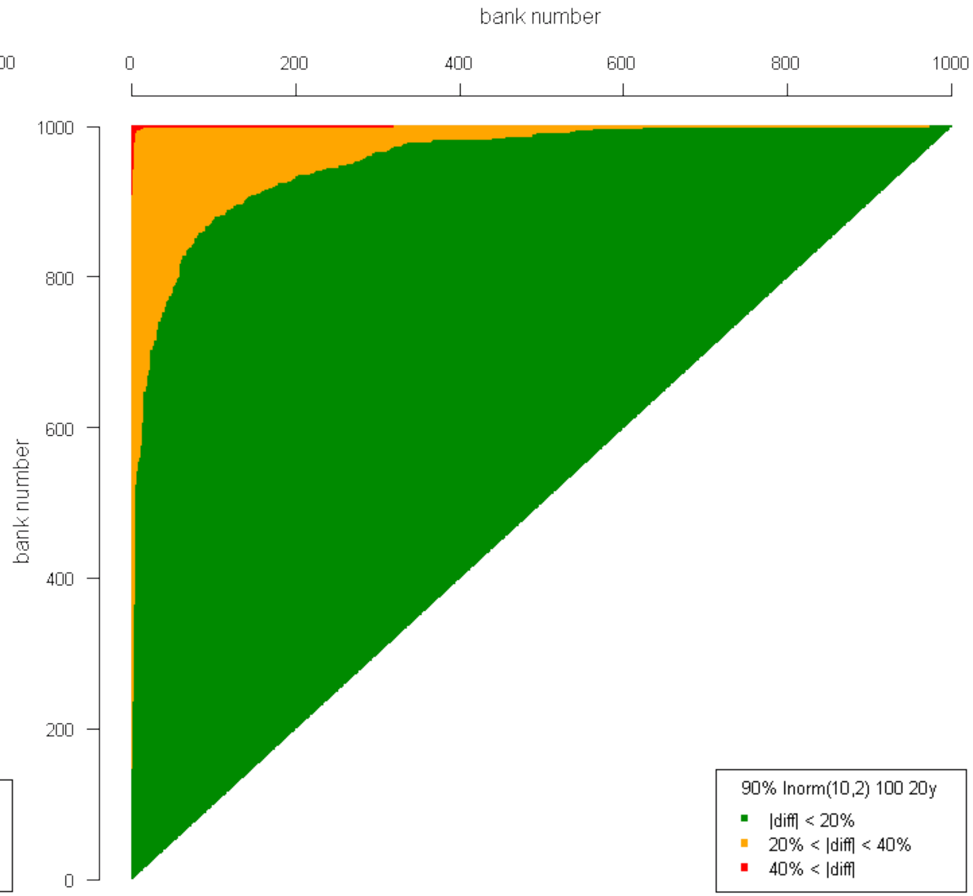
Representation of the pairwise difference of 99.9% VaR in % of the true value

Results – going “in sample” improves the situation

100 events per year – 5 years



100 events per year – 20 years



Representation of the pairwise difference of **90% VaR** in % of the true value

Results – a very unlevel playing field

- Even if two banks have the same risk profile in principle, their specific loss history makes it likely that one of them will have a much larger capital requirement than the other
  - Indeed, in the case study of moderate frequency, one in 10 banks will have a Capital Requirement larger than 1.35, and one in 10 smaller than 0.65: one has twice the CaR of the other, just by chance (is this fair?)
  - furthermore, the probability of a catastrophic loss (e.g. one producing a default) is the same for all 10 banks!
  - These numerical examples assumed a heavy, but not extreme, tail behaviour (lognormal, albeit with sigma=2) – these results will be exasperated with a heavier tail
  - When using EVT, the analysis is carried out in the tail, i.e. using a (generally much smaller) subsample, thus amplifying the statistical uncertainty and hence the level playing field problem
- The fraction of pairs with very large differences in VaR decreases when the VaR measure is closer to be "in sample", which can be achieved both by reducing the VaR confidence level and by using a longer loss history (i.e. a larger pool of banks)

LOGNORMAL SEVERITY - FREQ = 100

Loss history	VaR level	Green	Orange	Red
5 years	99.9%	41%	30%	29%
	90%	59%	30%	10%
20 years	99.9%	72%	25%	3%
	90%	91%	9%	0%

Practical consequences

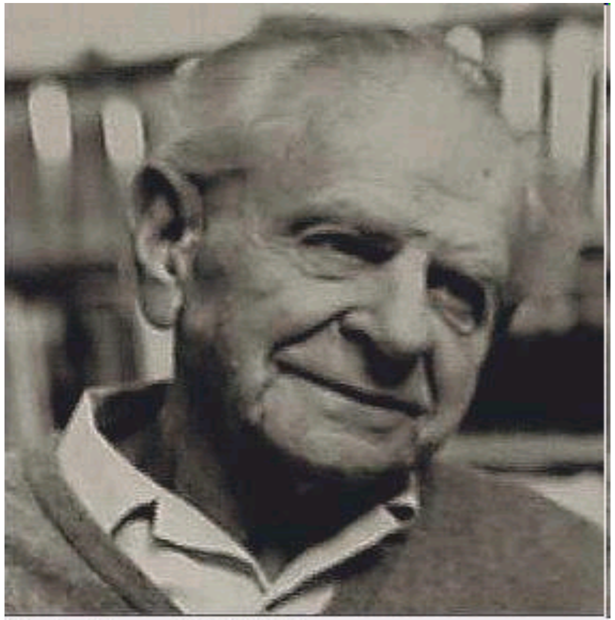
- Capital estimates at the required regulatory levels are highly uncertain
- Capital estimates are highly volatile from one period to the next with each new addition of extreme losses to the data sets

- Capital estimates and associated uncertainties can be driven by just one or two high-severity, low-frequency loss categories
- It is highly likely there will be an uneven – and unfair – distribution of capital levels across banks
- Even if models could theoretically produce results with a certain degree of precision, validation is practically impossible given the rate of change in banking practices

Can capital models be validated?

- Model validation requires that statements about loss probabilities are subjected to empirical testing
  - Can statements about capital adequacy be falsified?
    - Maybe, but only if they are very badly wrong
  - Following Popper, unfalsifiable statements are unscientific

- Suppose 1,000 bank-years worth of loss data are collected, then could models be validated?
  - Older data rapidly become irrelevant to predicting future losses
    - Mergers, banking trends, changes in risk management, changes in business practice, legal environment, etc.
  - Pooled data from several banks must therefore be used
    - Validation of the scaling models needed to pool data across diverse banks requires even more data
    - May require an external validation authority



Sir Karl Popper (1902-1994)

The need for quality data

- To calculate the likelihood of extreme events, it is critical that loss data be as high-quality as possible
  - Accurately measured and thoroughly reported
  - Consistently and correctly classified
  - Pooled across as many homogeneous sources as possible
  - Appropriately scaled across heterogeneous sources
- External data collected from voluntary public reports of losses tend to be highly biased
- Consortia such as ORX have rigorous data submission and quality control standards
  - Anonymity of records helps remove any incentive to withhold loss data
  - Definitions Working Groups help ensure consistency in labeling losses
  - Analytics Working Groups provide data quality control and outlier analysis

Conclusions

The financial industry has reached a high level of understanding of the Operational Risk discipline and a rather sophisticated level of modelling the risk profile

- Approaches to modelling operational risk losses have become well engineered with the use of smart statistical techniques
- Methods have been developed (in some cases intelligently borrowed from other disciplines) to solve difficult problems
- Software tools are available to “industrialize” the VaR production

… but we are still facing some “unpleasant” effects which are difficult to properly understand and to keep under control. In particular the unlucky definition of the risk measure imposed (confidence level and time horizon) by the prudential regulation (Basel II) is fundamentally flawing the soundness of the measurement framework. This has to:

- be understood in the first place
- worked around, even at the price of loosing some theoretical consistency
- trigger lobbying initiatives to properly change current regulation.

In order to bring the Operational Risk measurement framework back to a manageable situation, the regulation should be improved by

- **Giving a proper meaning to op risk measures and to model validation and, last but not least,**
- **Achieving the level playing field we all claim to want in place**

The way forward – a proposal for modesty

- There is probably no solution to these issues with the way in which the current regulation is formulated
- What can be done to alleviate these problems?
  - *Promote cross-bank loss sharing and benchmark models*. Consortia such as ORX are already addressing some of these issues. An external validation authority would also improve the soundness of banks' internal estimates of risk
  - *Incorporate structural knowledge of loss generation.* Further insight should be developed into the mechanisms that determine the severity of extreme losses, in conjunction with scenario analyses. Trends and other long-term changes in bank structure should be factored into predictive loss models
- Taking a more radical attitude
  - We should stop trying to measure the un-measurable and asking unanswerable questions. The regulatory standard encourages false confidence in the powers of risk quantification
  - Rather, a lower measurement standard, say around the 90-95th percentile on a quarterly time horizon, is attainable and could better guide risk management practice. Regulatory capital could be determined, e.g., as a fixed multiplier times the lower measurement standard as it is done, for instance, in the market risk regulation