Operational Risk Measurement and Management
2º Seminário Internacional sobre modelos avançados para risco operacional

Santiago Carrillo Menéndez

RiskLab-Madrid
Quantitative Risk Research

FEBRABAN
Sao Paulo June 2-5 2009
Introduction

- We suggest a way to AMA models in an integrated operational risk framework.
- Why advanced models for operational risk?
  - Answer may be very simple.
  - A first reason: because they imply a better knowledge of entity risk profile and a better management.
  - But also they imply a lower amount of capital.
    - May be not so important at the first moment.
    - But, in the standard model, capital is a linear function of gross incomes.
    - While empirical evidence seems to indicate that severity is not growing significantly in managed banks (in the opposite, it tends to decrease).
    - In addition, capital is far from being a linear function of frequency.
  - Even more, advanced models may imply a significant cost reduction:
    - in expected loss size and, for example,
    - in insurance premium due to the capability for demonstrate the real risk profile of the entity.
Contents

- Part I is dedicated to the **Loss Distribution Approach**.
  - Fundamental concepts and numerical procedures are explained.
  - Key words: severity; frequency; recursive method; Monte Carlo simulation; Fourier Transform methods.

- In Part II we study some important issues modeling operational risk as analytical formula, model risk, scenarios analysis and stress testing.
  - Robust implementing of operational risk measurement implies deep knowledge of several key issues.
  - Key words: analytical formulas; model risk; body effect; scenarios analysis; stress testing.

- The treatment of qualitative information, in particular bank specific business environment and internal control factors is not part of this presentation but we can speak about how they can also be used as inputs to the model to reflect changes that may affect the risk profile of the institution.
Part I

The Loss Distribution Approach
Introduction: Severity and Frequency

Introduction
Severity and Frequency
Stylized Facts About Operational Risk

Fitting Severity

Models for Frequency Distributions

Capital Calculation
Introduction

- LDA is a statistical/actuarial approach for computing aggregate loss distributions [Klugman et al. 1998].
- LDA comes from actuarial techniques as they have been developed by the insurance industry for years.
- Looking at operational risk modeling, it is the most natural idea.
- The fundamental premise underlying LDA is that for each firm, operational losses reflect its underlying operational risk exposure.
- Loss data is the most objective risk indicator currently available.
- Nevertheless, actuarial techniques could not be imported directly because of the specificities of operational risks:
  - reporting bias;
  - scarcity of data.
- These two features of operational risk data have a dramatic impact on capital charge and thus can not be neglected.
- It imposes to deal with more sophisticated computations than the one use in insurance models.
Severity and Frequency I

- Those are the results of the LDCE 2002 (QIS 3): 47,269 losses of more than €20,000, representing gross losses of €8 billions.

- Two random variables need to be modeled: severity and frequency.

<table>
<thead>
<tr>
<th>Severity and Frequency by business line</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of events (%)</td>
</tr>
<tr>
<td>-----------------------</td>
</tr>
<tr>
<td>Corporate finance</td>
</tr>
<tr>
<td>Trading &amp; sales</td>
</tr>
<tr>
<td>Retail banking</td>
</tr>
<tr>
<td>Commercial Banking</td>
</tr>
<tr>
<td>Payment &amp; settlement</td>
</tr>
<tr>
<td>Agency services</td>
</tr>
<tr>
<td>Asset management</td>
</tr>
<tr>
<td>Retail brokerage</td>
</tr>
</tbody>
</table>
Severity and Frequency II

<table>
<thead>
<tr>
<th></th>
<th>number of events (%)</th>
<th>gross losses (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal fraud</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>External Fraud</td>
<td>44</td>
<td>16</td>
</tr>
<tr>
<td>Employment</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>Clients, Products</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>Damage to</td>
<td>1</td>
<td>24</td>
</tr>
<tr>
<td>Business disruption</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Execution, delivery, ...</td>
<td>35</td>
<td>29</td>
</tr>
</tbody>
</table>

- The situation is very different through the different cells (or units).
- For each cell, we must fit a distribution for severity and another one for frequency and then combine both in order to produce the aggregate loss distribution of the cell.
- Two kind of events: low frequency-high impact, high frequency-low impact.
Stylized Facts About Operational Risk

- The number of loss events in a big European bank is huge ($\sim 100,000$ per year).
- Losses greater than €10,000 represent:
  - 50% of gross losses,
  - but less than 1% of the number of losses.
- Around 50% of the loss events represent losses smaller than €6-10.
- Operational losses appear to follow heavy-tailed distributions.
- More than 90% of the capital charge is usually explained by a very small number of events.
- Data points span many orders of magnitude.
- The largest loss is usually at 30 or more standard deviations away from the mean.
- Many authors suggest the use of Extreme Value Theory distributions in order to fit real data.
Introduction: Severity and Frequency

Fitting Severity
- Models for Severity Distributions
- More Complex Situations: the Real Word
- Graphical Tools for Goodness of Fit
- Using External Data
- Giving Weights to Data

Models for Frequency Distributions

Capital Calculation
# Models for Severity Distributions

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Expression</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lognormal</td>
<td>( H(x) = \mathcal{N}\left( \frac{\ln(x) - \mu}{\sigma} \right) )</td>
<td>( \mu, \sigma &gt; 0 )</td>
</tr>
<tr>
<td>GEV</td>
<td>( H(x) = \exp\left( - \left[ 1 + \xi \frac{x - \alpha}{\beta} \right]^{+ \frac{1}{\xi}} \right) )</td>
<td>( \alpha, \beta &gt; 0, \xi )</td>
</tr>
<tr>
<td>Pareto gen.</td>
<td>( H(x) = 1 - \left[ 1 + \xi \frac{x - \alpha}{\beta} \right]^{+ \frac{1}{\xi}} )</td>
<td>( \alpha, \beta &gt; 0, \xi )</td>
</tr>
<tr>
<td>Weibull</td>
<td>( H(x) = 1 - \exp\left( - \left[ \frac{x - \alpha}{\beta} \right]^{+ \xi} \right) )</td>
<td>( \alpha, \beta &gt; 0, \xi )</td>
</tr>
<tr>
<td>Gamma</td>
<td>( H'(x) = (x - \gamma)^{\alpha - 1} [\beta^\alpha \Gamma(\alpha)]^{-1} e^{-(x-\gamma)/\beta} )</td>
<td>( \alpha &gt; 0, \beta &gt; 0, \gamma &gt; 0 ),</td>
</tr>
<tr>
<td>Log-gamma</td>
<td>( X \sim \mathcal{LN}(\mu, \theta \times \sigma^2) )</td>
<td>( \theta \sim \Gamma(1/\beta, \beta) )</td>
</tr>
</tbody>
</table>

Inverse gaussian, Burr, g-and-h distributions are other possibilities.
The Real Dirty World

- In practice, we shall have more complicated situations.
- No single distribution fits well over the entire data set.

- So we need mixtures of our distributions.
Another typical situation occurs when body and tail need to be fitted separately.

In our case, the lognormal fitting gives a p-value 0.38 for Kolmogorov-Smirnov test (similar for Anderson-Därling).

When fitting separately, we get 0.55 for the body and 0.37 for the tail using a lognormal distribution and 0.45 for the Pareto.

For a single fitting of the data to Pareto distribution, the goodness of fit is 0.

For the body an empirical fitting may be a good solution.
Graphical Tools: PDF-Plot

- Of course, a first recommendation is to look at the goodness of fit plotting both empirical and theoretical pdf’s.
- An eventual “zoom” on the tail may be helpful.
Graphical Tools: QQ-Plot I

- QQ Plot represents the quantiles of the theoretical distribution (x-axis) against those of the model.
- If the model is good, points will lie very close to the line from 0 to 1.
- Several (theoretical) distributions may be tested at the same time in order to decide which gives the best fitting.
- the corresponding p-values are:

<table>
<thead>
<tr>
<th></th>
<th>Logn</th>
<th>Gam</th>
<th>Wei</th>
</tr>
</thead>
<tbody>
<tr>
<td>KS</td>
<td>0.88</td>
<td>0.06</td>
<td>0.00</td>
</tr>
<tr>
<td>AD</td>
<td>0.85</td>
<td>0.04</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Graphical Tools: QQ-Plot II

- We can use the QQ Plot for checking the goodness of fitting.
- Specially we can focus on tails.
- We generate 100 points normally distributed.
- Quantiles of the normal distribution are on the x-axis.
- Values obtained by formula on the y-axis.
Graphical Tools: Order Statistics I

- Let \( X_1, \ldots, X_N \) be iid rv with cdf \( F \). It is easy to see that

\[
G_N(x) = P(\max(X_1, \ldots, X_N) \leq x) = F(x)^N \Rightarrow g_N(x) = Nf(x)F(x)^{N-1}
\]

- This is the pdf of a **worst case scenario**.
- It allows for a tail discrimination between apparently similar distributions (normal and \( t \) Student with \( \nu = 3; N = 10 \)).
Graphical Tools: Order Statistics II

- Applied to data, we can get this kind of (good) situations.

![Severity Distribution Graph](image)
Using External Data I

- The use of external data is one of the main requirements of Basel II.

“Any risk measurement system must have certain key features to meet the supervisory soundness standard set out in this section. These elements must include the use of internal data, relevant external data, scenario analysis and factors reflecting the business environment and internal control systems.” (629)

“A bank operational risk measurement system must use relevant external data (either public data and/or pooled industry data), especially when there is reason to believe that the bank is exposed to infrequent, yet potentially severe, losses.” (634)

“These external data should include data on actual loss amounts, information on the scale of business operations where the event occurred, information on the causes and circumstances of the loss events or other information that would help in assessing the relevance of the loss event for other banks.” (634)

“A bank must have a systematic process for determining the situations for which external data must be used and the methodologies used to incorporate the data ...” (634)
Using External Data II

- The purpose of this requirement is to supplement the lack of high severity low frequency (quinquennial, decennial, ...) events in the internal database.

- In other words: internal severity databases should be supplemented with external severity data in order to give a non-zero likelihood to rare events which could be missing in internal databases.

- The most relevant aspects to be taken into account when deciding which external database to use are:
  - Quality and reliability of data (including quality control the way data are collected).
  - How representative are those data comparing with our entity.
  - In particular, are the data conveniently scaled and anonymous.
  - A well defined threshold is used for the data collection.
  - Availability over time.

- **ORX** (for Operational Riskdata eXchange) is an international, not for profit, consortium for sharing external data.
  - More than 51 members (February 2009) and growing.
  - 102,000 events, the biggest part with a threshold at €20,000.
  - Total gross losses bigger than €34 billions.
  - Regional and/or local subsets available.
Integrating Internal and External Data I

- As we outlined: external databases are strongly biased toward high-severity events.
- Mixing internal and external data altogether may then provide unacceptable results.
- A rigorous statistical treatment is required to make sure that merging both databases results in unbiased estimates of the severity distribution.
- **Fair Mixing Assumption**: external data are supposed to be drawn from the same distribution $f_\theta(x)$ as internal data except that external data are truncated above a threshold $H$.
- Under this assumption, external and internal data can be pooled together provided external data have been made comparable with internal data in order to increase the accuracy of estimators.
- **Known constant threshold assumption**: We suppose the threshold $H$ for external data is known.
- Let $X$ (resp. $X^*$) represent internal (external) data.
Integrating Internal and External Data II

- We have:

\[ X \sim f_\theta, \quad X^* \sim f_{\theta/H} \]

where \( f_{\theta/H} \) is the loss pdf conditionally to these losses being above threshold \( H \):

\[
f_{\theta/H}(x) = \frac{f_\theta(x)}{P(X > x)} 1\{x > H\} = \frac{f_\theta(x)}{\int_H^\infty f_\theta(u) du} 1\{x > H\}
\]

- Let \( X_j, j \in J \) (respect \( X_j^*, j \in J^* \)) denotes internal (resp. external) losses records.

- The maximum likelihood (ML) estimate is then the solution of the maximization problem:

\[
\hat{\theta} = \arg \max \left\{ \sum_{j \in J} \ln f_\theta(X_j) + \sum_{j \in J^*} \ln f_{\theta/H}(X_j^*) \right\}
\]

- This approach may be generalized in order to take into account an unknown or even a stochastic threshold [Baud et al. 2002].
Giving Weights to Data

- It may be interesting to give different weights to different data.
- For example, one may consider that data far in the past are less representative of the actual risk profile of the entity than recent data.
- A RiskMetrics like approach, based on exponentially weighted moving average (EWMA) is a possible solution for this problem.
- Given a sample $X_1, \ldots, X_N$ a weight $w_i$ is allocated to each observation $X_i$.
- The associated two first moments are then defined by

\[
\hat{\mu} = \sum_{i=1}^{N} w_i X_i, \quad \hat{\sigma}^2 = \sum_{i=1}^{N} w_i (X_i - \hat{\mu})^2
\]

- Similar formula may be produced for higher moments.
- The corresponding weighted likelihood function is the given by:

\[
\tilde{L}(\theta) = \sum_{i=1}^{N} w_i \ln[f_{\theta}(X_i)]
\]
Comparing Internal and External Data

- An interesting output of the use of external data is the possibility to measure the sensitivity of capital to external data.
- What is the variation of the capital using or not external data?
- For a judicious election of external data (for example the local or regional subset), interesting information may be extracted from this number.
- As we explained, classical statistical measures may not be the best tools.
- Variation of capital is the best benchmark for comparing internal and external data.
- Using the same threshold, both for internal and external data, we get a comparison with the sector.
- In addition, sensitivity measures may be driven varying the relative weight of the external database.
- Using similar techniques it would be possible to produce other interesting information like anonymous benchmarks (e.g. by business line and type of risk).
Configuring Internal Database

- Internal database is a key issue both for measurement and management of operational risk.
- Useful information for both purposes must be recollected.
- On one side,
  - Business lines and sub-lines, risk type and sub-risk types (for example the Basel II metrics).
  - Causes, consequences, . . . .
- On the other, the necessary information for fitting distributions: date and amount of the loss.
- Weights are necessary at least for two reasons:
  - In order to take into account the “real” frequency.
  - Last but not less, in order to take into account “real” severity in the case of great multi-cell loses
- Instead of splitting events in smaller losses across the corresponding cells, which would change the tail of the distribution, a better solution consists in the assignation of the total amount of the loss to each of the affected cells, but splitting the probability [Aue et al.].
Using Different Databases With Different Weights

- The likelihood function defined for the use of both internal and external databases may be extended to the use of as many databases as necessary.
- It allows, for example for the use of several internal data bases with different thresholds.
- In the same way, different external databases may be used.
- Even more, different weights may be assigned to different databases:
  - A bank may be more confident on its internal data than on external data: it can translate this view with a higher weight for internal data.
  - As well a bank may consider a regional or local subset of external data more representative of its risk profile than a global one and translate it in different weights.
  - For example, in the case of one internal and two external databases, the maximization problem would have the following expression:

\[
\hat{\theta} = \arg\max \left\{ w_1 \sum_{j \in J} \ln f_\theta(X_j) + w_2 \sum_{j \in J_1^*} \ln f_\theta/H_1(X_{1:j}^*) + w_3 \sum_{j \in J_2^*} \ln f_\theta/H_2(X_{2:j}^*) \right\}
\]

- Of course it is possible to combine different weights for different databases with different weights inside each of them.
Introduction: Severity and Frequency

Fitting Severity

Models for Frequency Distributions
  Introduction
  The Poisson Distribution
  Fitting Frequency to Data

Capital Calculation
Introduction

- Counting distributions are discrete distributions with probabilities only on natural numbers (including 0).
- In the operational risk framework, they will represent the frequencies of losses.
- First natural candidates are Poisson distributions.
- Nevertheless, because of the need for more flexibility in the shape the use of other distributions like mixtures of Poisson, negative binomial, and compound models is quite popular in operational risk.
- Other models like binomial or geometric distributions have no practical use in spite of their academical interest.
- We shall denote by Let $N$ be the frequency of losses in one cell and

$$p_k = P(N = k)$$

will represent the associated mass function.
The Poisson Distribution

- The **Poisson distribution** mass function (parameter $\lambda$ is defined by:

$$p_k = \frac{\lambda^k}{k!} e^{-\lambda}, \ \forall k \in \mathbb{N}$$

- The corresponding moment generating function is given by:

$$\phi(t) = \exp\{\lambda(e^t - 1)\}$$

which implies

$$E[N] = \lambda = \text{var}(N)$$

- For $N_1, N_2$ independent Poisson variables with parameters $\lambda_1, \lambda_2$, $X_1 + X_2 \sim \mathcal{P}(\lambda_1 + \lambda_2)$.

- This is a useful property for time aggregation.

- The parameter $\lambda$ is estimated as the empirical mean: $\hat{\lambda} = \bar{X}$. 

Santiago Carrillo Menéndez (RiskLab-QRR)
The Negative Binomial Distribution

- For the **Negative binomial distribution** the mass function can be written in two equivalent ways:

  \[
  p_k = \begin{cases} 
  \binom{k + r - 1}{r - 1} p^r (1 - p)^k, & p \in (0, 1); \\
  \binom{k + r - 1}{k} \left(\frac{1}{1 + \beta}\right)^r \left(\frac{\beta}{1 + \beta}\right)^k, & \beta > 0. 
  \end{cases} 
  \]

- The binomial coefficient is to be evaluated using:

  \[
  \binom{x}{k} = \frac{x(x - 1) \ldots (x - k + 1)}{k!} \quad \left(= \frac{\Gamma(x + 1)}{\Gamma(k + 1)\Gamma(x - k + 1)} \text{ if } x > k - 1 \right)
  \]

- The probability generating function is given by:

  \[P(z) = [1 - \beta(z - 1)]^{-r} \Rightarrow E[N] = r \beta, \quad \text{and} \quad \text{var}(N) = r \beta(1 + \beta).\]

- When observed variance is larger than the observed mean, the negative binomial is preferred to Poisson distribution for frequency fitting.

Santiago Carrillo Menéndez (RiskLab-QRR)
An Example

- Consider the following frequency table:

<table>
<thead>
<tr>
<th>Data Set 2</th>
<th>44</th>
<th>48</th>
<th>49</th>
<th>50</th>
<th>53</th>
<th>54</th>
<th>56</th>
<th>57</th>
<th>58</th>
<th>59</th>
<th>60</th>
<th>61</th>
<th>62</th>
<th>63</th>
<th>64</th>
<th>65</th>
<th>66</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_1$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N_2$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>68</td>
<td>69</td>
<td>70</td>
<td>71</td>
<td>72</td>
<td>73</td>
<td>74</td>
<td>75</td>
<td>76</td>
<td>77</td>
<td>78</td>
<td>79</td>
<td>82</td>
<td>91</td>
<td>93</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>$N_1$</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N_2$</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Empirical moments are:

$$\hat{\mu}_1 = 66.79, \quad \hat{\mu}_2 = 67.09, \quad \hat{\sigma}_1^2 = 57.40, \quad \hat{\sigma}_2^2 = 184.23$$

- This suggest a Poisson distribution for the first fitting and a negative binomial for the second one.

- The recursive expression allows for a graphical way to deciding the distribution to be fitted.

- In the case of a Poisson distribution, $kp_{k-1}/p_k$ must be more or less constant (equal to $\lambda$), while in the case of a negative binomial, it is an affine function.
About Poisson Distribution and Bias Collection

▶ A first idea: assume that the frequency distribution is Poisson.
▶ This distribution has many appealing features:
  1. There are strong arguments, supposing independence over time, in defense of a Poisson model.
  2. Because of this, it is widely used in the insurance industry for modeling problems similar to operational risks;
  3. Only one parameter is needed for the entire description and the ML value of this parameter is the empirical average number of events per year.
▶ However the way reporting is done may generate biases:
  ▶ If one bank is using a high truncating threshold, the average number of (reported) events will be low.
  ▶ If accounting is the origin of information, false seasonality will appear.
▶ In the first case, the average number of events must be corrected.
▶ In the second case, in spite of the fact that original data may be Poisson distributed, a negative binomial could give a much better fit.
▶ In practice, using negative binomial or Poisson may not suppose a big difference in the regulatory capital.
▶ Nevertheless it is important to demonstrate one is using the “right” model.
Taking into Account Truncation Thresholds

- The mentioned correction needs an estimate of the severity distribution for its calculation.
- Because of it, calibration of the frequency distribution comes after having calibrated the severity distribution.
- Let us consider first the case of a Poisson fitting: The “true” frequency parameter is given by

\[ \lambda = \frac{\lambda_{\text{sample}}}{1 - P(X > H)} \]

- In the case of negative binomial, the correction will also apply to the mean of the distribution \( E[N] = r\beta \):

\[ r = \frac{r_{\text{sample}}}{1 - P(X > H)} \]

- Parameters of the frequency distribution are estimated by maximum likelihood.

**Remark:** we shall not use external data for frequency fitting.
More About Thresholds I

- As we mentioned before, the use of truncating threshold may have serious impact on the selection of the severity distribution.
- But this fact can also impact deeply in the frequency distribution fitting.
- Consider the following example:

Without truncating threshold, we get:

- A mixture of lognormal and gamma distributions for the severity.
- A negative binomial \((r = 1424, p = 0.40, \text{mean } 2,136)\) for the frequency.
- Resulting capital is 164,815.
More About Thresholds II

- With a truncating threshold at 100, we get:
  - A Pareto distribution for the severity.
  - A negative binomial \( (r = 10707, p = 0.72, \text{mean } 4,183) \) for the frequency.
  - Resulting capital is 191,678.

- Scaling the frequency distribution to take into account the threshold has a big impact on frequency.
- The mean has been multiplied by a factor close to 2.
- Using the real probability on the left side of the threshold, this problem can be avoided.

Conclusions

- Data must be collected with a threshold as low as possible.
- It doesn’t mean 0-threshold is a requirement from the very beginning.
- Starting collecting and analyzing data is the important thing.
- ML allows for the fitting of data with different thresholds so threshold may be lowered later without losing the work previously done.
Introduction: Severity and Frequency

Fitting Severity

Models for Frequency Distributions

Capital Calculation
  The Aggregate Loss Distribution
  Monte Carlo Simulation
The Aggregate Loss Distribution

- Let $X$ be the severity of the losses (in a given unit), $N$ the frequency.
- The total aggregate loss is then:
  \[ S = \sum_{n=1}^{N} X_i, \quad X_i \sim X, \text{ i.i.d} \]
- It is a random sum of random variables.
- Moments are then easy to calculate:
  \[
  E[S_N] = E[X] E[N], \quad \text{var}[S_N^2] = E[N] \text{var}(X) + \text{var}(N) E[X]^2
  \]
- But not so the percentiles.
- The capital at risk (CaR) for this unit is defined as the percentile 99.9%.
- The total CaR is the sum of the quantities obtained in such a way.
- That means an unifactorial model (pessimistic).
- Alternative: dependence structure may be taken into account.
Monte Carlo Simulation I

- Monte Carlo simulation is another possibility for the description of the loss distribution.
Monte Carlo Simulation II

- For an accurate use of Monte Carlo simulation, several issues are to be taken into account.
- There is no arbitrage across numerical methods: for a large standard deviation in the severity and an important value for the frequency parameter a big amount of simulations will be necessary for a robust estimation of the percentile 99.9.
- In general, particularly for taking into account dependence structure, a huge amount of simulations may be needed.
- It is necessary to have at disposal robust random numbers generator(s).
- The precision of the output must be controlled: it is not sufficient to decide the number of simulations ex ante.
- Control variables (first moments of the aggregate loss) could be useful.
- Panjer’s recursive formula and Fourier Transform are other possibilities for numerical calculation.
Part II

Some Important Issues
Model Risk

The Percentile 99.9
Threshold Effect
Instability of Pareto fitting
The Body Effect

Analytical Approximations

Goodness of Fit Revisited

Risk Mitigation

Scenarios Analysis and Stress Testing
The Percentile 99.9

- One of the biggest issues in the LDA approach is the accurate calculation of the aggregate loss distribution 99.9 percentile.
- This requires a precise fit of the tail of the severity distribution.
- In practice, different distributions may offer similar goodness of fit to data with very different results in terms of capital:
  - different lognormals with high sigma,
  - g-and-h distributions,
  - Pareto distribution.
- The goal is to extrapolate the shape of the severity distribution far in the tail, based on the knowledge of part of the body.
- It may be very difficult (lack of sufficient data far in the tail) to distinguish between them.
- Model error is a real threat.
- Especially when high thresholds are used.
Severity Uncertainty I

- Data are sparse in the tails.
- There may not be enough empirical evidence to select model distributions with very different asymptotic behavior.
- Following [Carrillo-Suárez 2007], let us consider, for example:
  - a lognormal ($\mu = 10$, $\sigma = 2.5$), the histogram;
  - a piecewise defined distribution with a lognormal body and a $g$-and-$h$ ($a = 0.5$, $b = 5 \times 10^4$, $g = 2.25$ and $h = 0.25$) tail (15% of data, $u_0 = 3 \times 10^5$);
  - a piecewise defined distribution with a lognormal body and a generalized Pareto ($u = u_0$, $\beta = 5 \times 10^5$, $\xi = 1$) tail (15% of data, $u_0 = 3 \times 10^5$).

- We are considering really heavy tailed distributions.
- In the following figures we compare the lognormal distribution (histogram) with the lognormal + $h$-and-$g$ (left) and lognormal + Pareto (right).
Severity Uncertainty II

- The tail profiles of these distributions are very similar except very far in the tails.

- The asymptotic behaviors of the distributions are very different.

- Thus the CaR (or OpVar) associated to those distributions ($\lambda = 200$) are, respectively:
  - $1.42 \times 10^9$ (lognormal),
  - $6.21 \times 10^9$ (g-and-h) and
  - $1.54 \times 10^{10}$ (generalized Pareto),
Threshold Effect

- To illustrate the effect of varying the left threshold on the error in the model, we generate 30,000 lognormal random numbers for different values of $\sigma (\mu = 0)$.
- Two threshold levels are chosen (6,000 and 12,000).
- We present the results of the best fit (worst p-value of KS and AD tests).

<table>
<thead>
<tr>
<th>Threshold</th>
<th>6,000</th>
<th>10,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.75</td>
<td>Pareto</td>
<td>Pareto</td>
</tr>
<tr>
<td>1.00</td>
<td>Weibull</td>
<td>Weibull</td>
</tr>
<tr>
<td>1.25</td>
<td>Pareto</td>
<td>Pareto</td>
</tr>
<tr>
<td>1.50</td>
<td>Lognormal</td>
<td>GEV</td>
</tr>
<tr>
<td>1.75</td>
<td>Lognormal</td>
<td>Lognormal</td>
</tr>
<tr>
<td>2.00</td>
<td>Weibull</td>
<td>Lognormal</td>
</tr>
<tr>
<td>2.25</td>
<td>Lognormal</td>
<td>Pareto</td>
</tr>
<tr>
<td>2.50</td>
<td>Lognormal</td>
<td>Lognormal</td>
</tr>
</tbody>
</table>

- What's about less well defined data.
Impact on Capital

- The impact on capital depends on the frequency of events.
- The frequency distribution must be corrected in order to take into account the probability mass of the losses under the left censoring threshold.
- For high frequencies, the impact on capital may be very important:

<table>
<thead>
<tr>
<th>Threshold</th>
<th>6,000</th>
<th>10,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>σ</td>
<td>capital</td>
<td>variation</td>
</tr>
<tr>
<td>0.75</td>
<td>24,722,851</td>
<td>-45.14%</td>
</tr>
<tr>
<td>1.00</td>
<td>61,776,662</td>
<td>11.89%</td>
</tr>
<tr>
<td>1.25</td>
<td>64,931,049</td>
<td>-13.72%</td>
</tr>
<tr>
<td>1.50</td>
<td>114,193,654</td>
<td>6.92%</td>
</tr>
<tr>
<td>2.00</td>
<td>263,774,070</td>
<td>-11.75%</td>
</tr>
<tr>
<td>2.25</td>
<td>677,462,871</td>
<td>11.04%</td>
</tr>
<tr>
<td>2.50</td>
<td>1,825,327,187</td>
<td>14.31%</td>
</tr>
</tbody>
</table>
Critical Issues when Using Pareto Distribution (I) 

\( \xi = 0.6 \)

- When fitting Pareto to actual loss data, it is usual to get high values for \( \xi \) (even greater than 1).
- The parameters estimates in the Pareto fit are unstable.
- The (absolute) fluctuations of economic capital are very important.
- In order to illustrate this, we generate (Pareto, \( \xi = 0.6 \)) 30 events greater than €10,000 quarterly and fit data to Pareto at the end of each period.
- It doesn’t seem to be an acceptable solution.
Critical Issues when Using Pareto Distribution (II)

$\xi = 1.1$

- For $\xi > 1$, the situation is even more dramatic.
- First, the expected value of the losses is infinite. Therefore one should expect problems of consistency in the calculation of economic capital.
- With real data it is very easy to get extremely unrealistic amounts of capital.
- In general, Pareto fits tend to overestimate the value of capital-at-risk.
The Body Effect I

- Should Single-Loss Events Determine the Economic Capital?
- In a subexponential framework, high percentiles of the loss distributions levels are explained by a single extreme loss or a small amount of large losses.
- The value 99.9% is a very high percentile.
- Is it high enough in order to make the body (the part under the threshold) of the distribution irrelevant for the CaR calculation?
- In order to give an answer to this question, we perform the following simulation [Carrillo-Suárez 2007].
  - Random values of the loss severity are generated with a lognormal distribution ($\mu = 5, \sigma = 2$).
  - Different thresholds ($u$), determined by the probability of the tail ($p$), are chosen.
  - Three cases are considerate:
    1. case 0: empirical data;
    2. case 1: the losses under the threshold $u$ are all equal to 0;
    3. case 2: the losses under the threshold $u$ are all equal to $u$;
The results for the CaR (in thousands) are displayed in the following table.

In the case of conditional CaR the tendencies are similar.

<table>
<thead>
<tr>
<th></th>
<th>p</th>
<th>u</th>
<th>VaR₀</th>
<th>VaR₁</th>
<th>VaR₂</th>
<th>(\frac{\text{VaR}_2 - \text{VaR}_1}{\text{VaR}_0})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\lambda = 200)</td>
<td>0.50</td>
<td>149</td>
<td>1,253</td>
<td>1,249</td>
<td>1,263</td>
<td>1,14%</td>
</tr>
<tr>
<td></td>
<td>0.20</td>
<td>797</td>
<td>1,249</td>
<td>1,223</td>
<td>1,349</td>
<td>10.09%</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>1,946</td>
<td>1,256</td>
<td>1,204</td>
<td>1,557</td>
<td>28.05%</td>
</tr>
<tr>
<td>(\lambda = 2,000)</td>
<td>0.50</td>
<td>149</td>
<td>4,909</td>
<td>4,858</td>
<td>5,010</td>
<td>3.10%</td>
</tr>
<tr>
<td></td>
<td>0.20</td>
<td>797</td>
<td>4,896</td>
<td>4,624</td>
<td>5,897</td>
<td>25.98%</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>1,946</td>
<td>4,911</td>
<td>4,399</td>
<td>7,903</td>
<td>71.36%</td>
</tr>
<tr>
<td>(\lambda = 20,000)</td>
<td>0.50</td>
<td>149</td>
<td>28,655</td>
<td>28,126</td>
<td>29,653</td>
<td>5.32%</td>
</tr>
<tr>
<td></td>
<td>0.20</td>
<td>797</td>
<td>28,567</td>
<td>25,853</td>
<td>38,660</td>
<td>44.83%</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>1,946</td>
<td>28,727</td>
<td>23,519</td>
<td>58,630</td>
<td>122.22%</td>
</tr>
</tbody>
</table>
Conclusions:

- For low frequencies, the contribution of the body of the distribution is not decisive.
- Nevertheless the greater is the frequency the larger is the contribution of the body of the distribution.
- The asymptotic approximation works well for small frequencies.
- Note however that, in these experiments, the probability mass of the losses under the threshold is the same in all cases.
- If the probability mass of the body is extrapolated from the tail fit, these figures would show a much larger variation.
- From a practical point of view this suggests to collect data with a low truncating threshold.
- This point of view may have important returns in the management of the expected loss and insurance premiums negotiation.
- That means *costs reduction*. 
Model Risk

Analytical Approximations
- Subexponential Distributions
- An Analytical Formula
- Verifying the Böcker Klüpelberg Formula

Goodness of Fit Revisited

Risk Mitigation

Scenarios Analysis and Stress Testing
Subexponential Distributions

- Consider $X_1, \ldots, X_n, \ldots$, independent, identically distributed random variables with distribution function $F = F_X$ ($X_i \sim X, \forall i$).
- They belong to the class of subexponential distributions iff we have

$$\lim_{x \to \infty} \frac{P(X_1 + \cdots + X_n > x)}{P(\max(X_1, \ldots, X_n) > x)} = 1 \quad \text{for some (all) } n \geq 2$$

- This means that severe overall losses are mainly due to a single large loss rather than the consequence of accumulated small independent losses.
- It can be shown that this equation is equivalent to:

$$\lim_{x \to \infty} \frac{F^{*n}(x)}{F(x)} = n \quad \text{for some (all) } n \geq 2 \quad (\bar{H}(x) = 1 - H(x))$$

- A consequence is that, if the severity distribution $F$ is subexponential and

$$\sum_{n=0}^{\infty} (1 + \epsilon)^n P(N = n) < \infty, \text{ for some } \epsilon > 0$$

then $S$ is subexponential and

$$F_S(x) \sim E[N] \times F(x), \quad x \to \infty$$
An Analytical Formula (Böcker-Klüppelberg)

- Let us suppose \( F_S(x) \sim E[N] \bar{F}(x) \).
- If \( x_\kappa \) is such that \( F_S(x_\kappa) = \kappa \), we shall have (single-loss approximation):
  \[
  \text{VaR}_\kappa(F_S) = x_\kappa \sim F^{-1}\left(1 - \frac{1 - \kappa}{E[N]}\right)
  \] (1)

- Assuming we are in the asymptotic regime, we can compute the OpVaR.
- For example, for \( E[N] = 100 \), we get that \( \text{VaR}_{99.9\%}(F_S) \sim \text{VaR}_{99.999\%}(F) \).
- The use of an additional mean correction term improve the result.

An example: the Pareto case:

- Let \( F_U \) be the peak over threshold (POT) distribution of the severity:
  \[
  F_U(x) = P(X-u \leq x/X > u) \quad \forall 0 \leq x < x_F-u, x_F = \sup\{x > 0/F(x) < 1\}
  \]

- An elementary calculation leads to \( \bar{F}(x) = \bar{F}(u) \times \bar{F}_u(x-u) \)

- Using the Pareto approximation to the POT distribution (Balkema theorem), we get (for \( \kappa \to 1 \)):
  \[
  \bar{F}(x) \sim \bar{F}(u) \times \left(1 + \frac{\xi (x-u)}{\beta}\right)^{-1/\xi} \Rightarrow \text{VaR}_\kappa(G) \sim u + \frac{\beta}{\xi} \left[\left(\frac{\bar{F}(u) E[N]}{1 - \kappa}\right)^{\xi} - 1\right]
  \]
Verifying Böcker Klüpelberg Formula I

- We need to contrast the goodness of the Böcker-Klüpelberg formula.
- In order to complete the analysis, let us observe that if the frequency is a geometrical r.v. of parameter \( p \), Rényi theorem gives:

\[
P(S_N \leq \frac{x}{p}) \to 1 - e^{-x/\mu}
\]

- If we consider now the case of a negative binomial \((r,p)\), we obtain:

\[
P(S_N \leq \frac{x}{p}) \to 1 - \Gamma_{r,1/\mu}(x)
\]

where \( \Gamma_{r,1/\mu} \) is the probability distribution function of the gamma distribution of parameters \((r, 1/\mu)\).

- In the following table, we compare the Bócker-Klüppelberg formula (B-K) with “fair” (Monte Carlo) value of the operational VaR (in thousand of euros) for different values of the parameters \( p \) and \( r \).
### Verifying Böcker Klüpelberg Formula II

<table>
<thead>
<tr>
<th>p</th>
<th>r</th>
<th>E[N]</th>
<th>B-K</th>
<th>Rényi</th>
<th>MC</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>10</td>
<td>10</td>
<td>252</td>
<td>50</td>
<td>255±37</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>100</td>
<td>751</td>
<td>293</td>
<td>902±109</td>
</tr>
<tr>
<td></td>
<td>1,000</td>
<td>1,000</td>
<td>1,996</td>
<td>2,414</td>
<td>3,144±205</td>
</tr>
<tr>
<td></td>
<td>10,000</td>
<td>10,000</td>
<td>4,870</td>
<td>22,617</td>
<td>16,017±607</td>
</tr>
<tr>
<td>0.05</td>
<td>0.5263</td>
<td>10</td>
<td>252</td>
<td>121</td>
<td>292±42</td>
</tr>
<tr>
<td></td>
<td>5.2632</td>
<td>100</td>
<td>751</td>
<td>334</td>
<td>888±108</td>
</tr>
<tr>
<td></td>
<td>52.6316</td>
<td>1,000</td>
<td>1,996</td>
<td>1,709</td>
<td>3,218±301</td>
</tr>
<tr>
<td></td>
<td>526.3158</td>
<td>10,000</td>
<td>4,870</td>
<td>13,161</td>
<td>16,284±396</td>
</tr>
<tr>
<td>0.005</td>
<td>0.0503</td>
<td>10</td>
<td>252</td>
<td>601</td>
<td>668±60</td>
</tr>
<tr>
<td></td>
<td>0.5025</td>
<td>100</td>
<td>751</td>
<td>1,189</td>
<td>1,417±128</td>
</tr>
<tr>
<td></td>
<td>5.0251</td>
<td>1,000</td>
<td>1,996</td>
<td>3,254</td>
<td>3,871±191</td>
</tr>
<tr>
<td></td>
<td>50.2513</td>
<td>10,000</td>
<td>4,870</td>
<td>16,456</td>
<td>17,890±375</td>
</tr>
</tbody>
</table>

- The single-loss formula seems to work fine only for a small number of events.
Model Risk

Analytical Approximations

Goodness of Fit Revisited
  Giving Economic Sense to Capital
  Robustness of Fitting
  Taking into Account Qualitative Inputs
  Backtesting

Risk Mitigation

Scenarios Analysis and Stress Testing
Giving Economic Sense to Capital

- As we outlined in the case of Pareto distributions, in operational risk modeling it is not unusual to get models which reasonable goodness of fit but which make no sense from an economical point of view.
- What matters the goodness of fit if your regulatory capital has the same order of magnitude of the GNP of your country and your bank is well managed?
- Is a model acceptable if it produces great fluctuations of capital, quarter after quarter without perceptible changes in the risk profile of your entity?
- Is it reasonable to use a model which is very sensible to new data (outliers)?
- What is the economic sense of a model which places high probabilities to unrealistic size of losses?
- With some types of distributions it is relatively easy to produce this kind of situations.
- For example using distributions with infinite variance or even infinite mean.
- A critical analysis of choices made for severity distribution modeling must me done and compared with the real risk profile of the entity.
Robustness of Fitting

- This is one of the key issues in severity modeling.
- Fluctuations in the capital calculation, produced by a deficient model are unjustifiable.
- Robustness is a statistical property, but also an economical one.
- A first question is:
  - How robust is the model?
  - Does it change quarter after quarter (f.e. from Pareto to lognormal or Weibull)
- A second question is:
  - How stable are the parameters?,
  - and, more important, how sensible is the CaR to those parameters?
- A constant monitoring of those aspects is necessary.
- For example, a simple test consists in comparing the actual capital with the one we get by:
  - Doubling the frequency of the higher loss.
  - Picking it out from the database.
Taking into Account Qualitative Inputs

- The principal criticism to the methodology exposed here is to be backward-looking (based on passed data).
- How to make the capital calculation more forward-looking and how to reflect the bank’s risk controls and operating profile?
- Taking into account bank-specific business environment and internal control factors as required by Basel II.
- From such a qualitative analysis we can derive inputs for our quantitative model for capital calculation.
- For example, a better management may lead to the view that
  - Frequency has been lowered by a 10%.
  - Severity has been reduced, in mean, by a 15%.
  - A cap has been activated for a certain type of losses.
- All those views may be translated into the model in order to produce capital calculation.
Backtesting

- Backtesting is the sequential testing of a model against reality to check the accuracy of the predictions.
- Backtesting is a requirement for market risk: daily VaR measurement are compared with the day after variations in order to validate the accuracy of the model.
- In credit risk and operational risk, this approach to backtesting is no feasible because of the lack of (yearly) data.
- Other approaches are necessary.
- A first benchmark is, of course, capital requirement under TSA model.
- In a single loss event framework a comparison between percentiles of the aggregate loss distribution and shifted percentiles of the severity loss is a first possibility [Aue et al.].
- Another possible test is to check the likelihood of the number of the $n$ bigger losses.
Model Risk

Analytical Approximations

Goodness of Fit Revisited

Risk Mitigation
  Insurance and Operational Risk

Scenarios Analysis and Stress Testing
Insurance and Operational Risk I

- Insurance is a well-established risk management tool that has been used by the banking sector for many time.
- Basel II requires a sound methodology for calculating the capital reduction resulting from insurance.
- In addition, there is a cap of 20% in the total reduction at the bank level.
- Because of this, the real effect of insurance is a global effect in the entity which cannot been merely seeing as the sum of the effects on each cell of the Basel II array.
- As a consequence, the treatment of Insurances and the risk mitigation resulting must be driven through Monte Carlo simulation at the entity level.
- A real problem is, up to date, the lack of actuarial products tailored to fit the necessities of operational risk management.
- Nevertheless it may be useful to be able to study the impact on each cell (unit) of a concrete insurance.
- In particular for negotiating the price of insurances.
Insurance contracts are characterized by certain parameters:

- A prime $P$: it is the cost of the insurance contract.
- A deductible $d$ is defined to be the amount the bank has to cover by itself.
- The single limit $l$ of an insurance policy determines the maximum amount of a single loss that is compensated by the insurer.
- The probability of coverage $p$ describes the probability of application of the insurance contract. It may be used to model its maturity.
- A recovery rate $r$.

With those definition, the recovery for a given loss $x$ would be:

$$R_x = \begin{cases} 0, & \text{with probability } p; \\ r[\min(x, l) - d], & \text{with probability } 1 - p. \end{cases}$$

It is possible to reflect this parameters in the loss distribution and calculate (FFT, Monte Carlo) the effect of insurance.

In addition, an aggregate limit (a cap for the total exposure of the insurance company) may exist.
Model Risk

Analytical Approximations

Goodness of Fit Revisited

Risk Mitigation

Scenarios Analysis and Stress Testing
  Scenario Analysis
  Stress Testing
A Consistent Framework for Scenario Analysis I

- A typical LDA framework for capital calculation may involve the use of several databases.

- The internal database:

\[ D_i, \text{ weight } w_i, \text{ eventual threshold } H_i. \]

will be the core for all the statistical procedure.

- For some units (f.e. cells of the Basel II array), more at the first moments (years), a lack of data is a realistic hypothesis.

- This fail may be corrected by the use of complementary databases made of synthetical data (scenarios), based on expert's opinion:

\[ D_c:1, \ldots, D_c:n_c, \text{ with weights } w_c:1, \ldots, w_c:n_c. \]

- Several external databases may be used:

\[ D_e:1, \ldots, D_e:n_e, \text{ with weights } w_e:1, \ldots, w_e:n_e, \text{ and thresholds } H_e:1, \ldots, H_e:n_e. \]

- In addition to ORX global and regional or local data sets, databases of public events (f.e. OpVantage) or data from other banks of the group or pool may be used.
A Consistent Framework for Scenario Analysis II

- Now a scenario (for severity) is simply a vector of weights
  \[ W = (w_i, w_{c:1}, \ldots, w_{c:n_1}, w_{e:1}, \ldots, w_{e:n_e}) \]
  which represent the contribution of each database to the risk profile of the entity.
- The weights are based on expert’s opinions or systematical research.
- For each scenario the CaR is calculated based on the same methodology used for economic or regulatory capital.
- Results may be compared.
- It is only a question of computational recourses and they are cheap (grid computing).
- This is a **consistent way for severity scenario analysis**.
- Scenarios for severity may be combined with scenarios for frequency.
- Splitting both contributions allows for a more granular study of contributions to capital.
Stress Testing

- For **stress testing** purposes, we need to define new stress databases.
  
  \[ D_{s:1}, \ldots, D_{s:n_s}, \text{ with weights: } w_{s:1}, \ldots, w_{s:n_s}. \]

- Those databases are fulfilled by synthetical events, based on expert’s opinion, which represents extreme losses which could happen.

- They are worst case scenarios.

- Each of them must have a different weight in the database depending on its (supposed) frequency.

- For example, an hypothetical event which may happen once each 20 years would have a weight 0.5 if the database benchmark are decennial events.

- **Observation**: not taking into account different frequencies for extreme events would be equivalent to suppose correlation 1 for all the stress events, which is totally unrealistic.

- In our framework, a stress scenario a scenario with weight(s) different from zero for stress databases.

- It can include also views about dependence structure.
Part III

Bibliography
Bibliography I

Cruz M.G.  
Modelling Measuring and Hedging Operational Risk.  
*John Wiley & Sons, 2002.*

Klugman S.A., Panjer H.H., Willmot G.E.  
Loss Models, From Data to Decisions.  
*John Wiley & Sons. 1998.*

McNeil A.J., Frey R. & Embrecht P.  
Quantitative Risk Management.  
*Princeton Series in Finance, 2005.*

Panjer H.H.  
Operational Risk: Modeling Analytics.  

Panjer, H. & Willmot G.  
Insurance Risk Models.  
*Society of Actuaries Chicago 1992.*
Roncalli T.
La Gestión des Risques Financiers.

Aue, F. & Kalkbrener, M.
LDA at Work: Deutsche Bank’s Approach to Quantifying Operational Risk.

Baud, N. Frachot, A. & Roncalli T.
Internal data, external data and consortium data for operational risk measurement: How to pool data properly?

Basel Committee
February 2003.
Bibliography III

Carrillo Menéndez S. & Suárez A.  
Medición efectiva del riesgo operacional.  
*Estabilidad Financiera*, November 2006.

Carrillo Menéndez, S. & Suárez, A.  
Robust quantification of the exposure to operational risk: Bringing economic sense to economic capital.  

Ferreras Salagre A.  
Riesgo Operacional: Algunas consideraciones críticas relativas al uso de modelos avanzados.  

Mignola, G. & Ugoccioni, R.  
Sources of uncertainty in modelling operational risk losses.  
Moscadelli M.
The Modelling of Operational Risk: Experience with the Analysis of the Data Collected by the Basel Committee.
*Banca D’Italia, Termini di discussione N° 517.*

Scandizzo S.
Risk Mapping and Key Risk Indicators in Operational Risk Management.