Caserta, 14 march 2007

# Merging data streams in operational risk management


Paolo Giudici

**Department of Statistics "L. Lenti"**
**University of Pavia**

giudici@unipv.it


1

## Agenda

- Background
- Current approaches
- Our proposed approach
- Open issues

2

# Merging data streams in operational risk management

## Background



## Statistical models for *risk management*

- **Legislative background:** regulations of Banking Authorities (Basel2 regulations, www.bis.org); but also information security (ISO 17799) and business continuity (BS 25999) require risk measurement
- **Strategic background** is to improve controls, efficiency and performance of (service) companies

## Statistical models for *financial risk management*

**Objectives of modelling:**

- efficient capital allocation**,** to cover unexpected losses without allocating too much capital
- effective performance monitoring of business units, branches and processes

**Necessary condition** is a valid measurement of risks: good data quality, good data mining

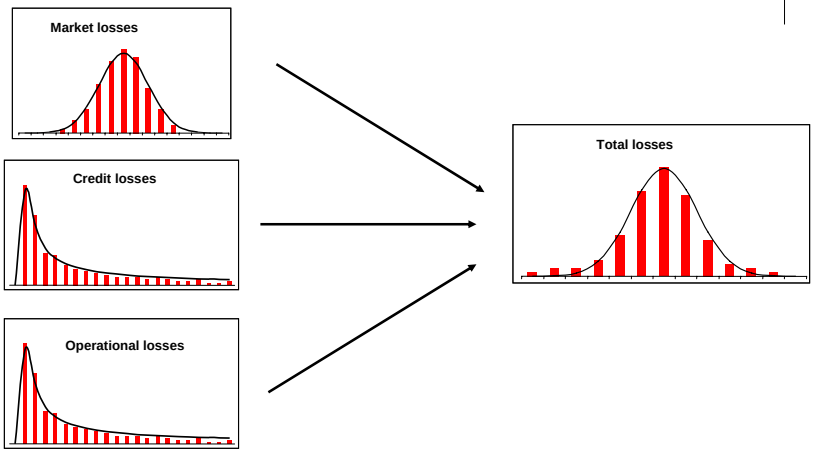
5

## *An (integrated) statistical view on risk management*

- Objective: to estimate a loss distribution, and derive functionals of interest from it (such as the Value at Risk)
- Losses in market risk are realisations of a continuous stochastic process
- Losses in credit risk are realisations of a convolution between a binary process (default or not) and a continuous one
- Losses in operational risk are realisations of a convolution between a counting process (frequency) and a number of continuous ones (severities)

6

# Final aim: distribution of integrated losses




7

# Operational risk

- Risk of financial losses caused by: inadequate internal processes, human errors, IT failures or external events. (Basel Committee, *Working Paper on the Regulatory Treatment of Operational Risk*, September 2001).
- Measures risks due to enterprise management (rather than financial or credit management).
- Compulsory for most banks by end of 2007

8

## *Advanced Measurement Approaches*

- Suited for large sized and efficient banks, and those operating at the international level
- Based on the analysis of all available and relevant data, by means of a statistical model aimed at estimating the probability distribution of the losses
- Most used statistical models are: scorecard, actuarial. Also causal

9

## Merging data streams in operational risk management

## Current approaches

10

## *Scorecard models*

- Based on expert opinions (of process owners) on operational losses expected for a future period (e.g. next year)
- Each process owner evaluates: which are the likely risks; the expected losses; possible causal factors (key risk indicators) of the losses
- In the most diffuse version assessments are in terms of expected frequencies and severities. Typically both on a categorical scale (low, medium,high).

11

## Actuarial models

- Employs two types of probability distributions:
- the distribution of the *frequency* of the risk events
- the distribution of the losses that arise for each given event (*severity*).
- The convolution between the two determines the loss distributions and, then, the Value at Risk.
- Suited for rare events, with high severity. Do not model relationships between loss events and with their causes. Requires simulation based methods to derive the loss distribution.

12

## Actuarial methods

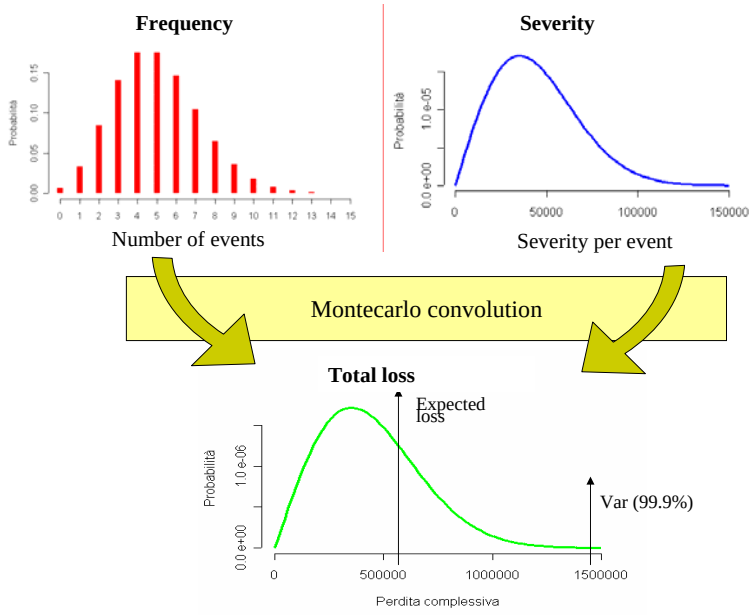
- For each risk type, business unit, and a given period, operational risk losses can be defined as a sum ($S$) of a random number ($n$) of single losses ($X_i$):

$$S = X_1 + X_2 + \ldots + X_n$$

- The actuarial model typically assumes that: *a)* single losses are i.i.d. random variables; *b)* the distribution of $N$ (frequency) is independent on that of $X_i$ (severity)

13

## Actuarial models



## Monte Carlo simulation

| RUN # | FREQ. | SEVERITY | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 First Event | | 2 Second event | | 3 Third event | |
| 1 | 0 | | | | | | |
| 2 | 3 | 0.247963 | 94.10739958 | 0.984985 | 300744623.6 | 0.227363 | 0.227086103 |
| 3 | 1 | 0.773522 | 25007.66528 | | | | |
| 4 | 0 | | | | | | |
| 5 | 0 | | | | | | |
| 6 | 0 | | | | | | |
| 7 | 1 | 0.982238 | 4872414.65 | | | | |
| 8 | 1 | 0.770287 | 23985.12518 | | | | |
| 9 | 1 | 0.287759 | 150.8397754 | | | | |
| 10 | 3 | 0.027833 | 0.768589688 | 0.041871 | 0.000193265 | 0.374248 | 4.911219919 |
| 11 | 0 | | | | | | |
| 12 | 3 | 0.938475 | 548073.9994 | 0.933653 | 2484677.895 | 0.615192 | 406.9620162 |
| (.....) | (.....) | | | | | | |
| 10000 | (.....) | | | | | | |

15

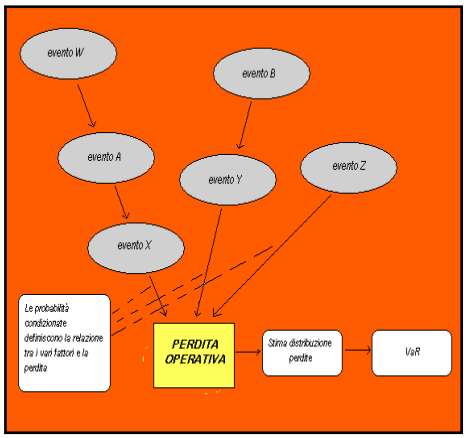
## Bayesian actuarial models

- Can add a prior distribution to both the frequency and severity distribution.
- A conjugate and little informative approach can be chosen.
- E.g. gamma prior for both the frequency and the severity distribution

16

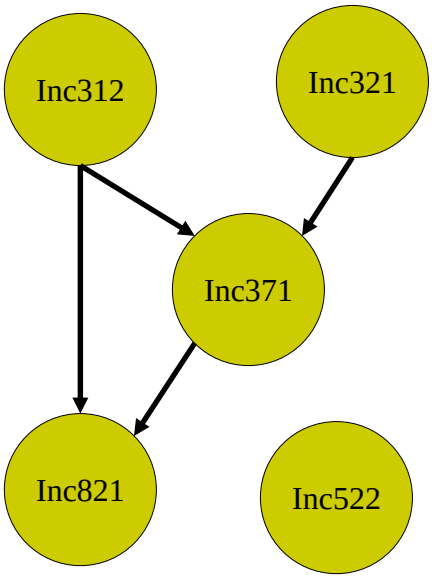
## Causal models

- Each operational loss depends on the outcome of other events (causes), each of which can also be dependent
- Suited for frequent events, with low severity
- Difficult to specify, yet very useful for audit and control purposes




17

## Bayesian networks




Advantages:

1) Can exploit correlations and causations with all risk factors
2) Can handle fusion of different sources of information;

Disadvantages:

1) Joint treatment of the two components of the expected loss (frequency and severity);

18

# Merging data streams in operational risk management

## Our proposed approach



# Basel 2 requirements

From the official documentation (www.bis.org) it can be desumed that AMA models must be based on the combined usage of four data sources:

"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**"

e.g. A sound AMA approach must combine**: historical loss data, external (consortium) data, expert opinions, key risk indicator data**

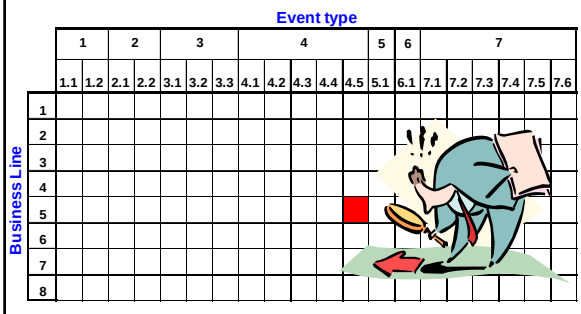
Our proposal aims to fulfill this requirement**.**

# How to integrate ?

- The open problem is how to integrate, in a statistically coherent and computationally efficient way, different sources of data.
- We have proposed an approach based on: non parametric models to prioritise risks and calculate the value at risk. This will be described here.
- A different development involves lifting the assumption of perfect dependence among risk types, using Bayesian networks and Copula models. See e.g. Bonafede and Giudici (2007) and Dalla Valle, Giudici and Fantazzini (2007)



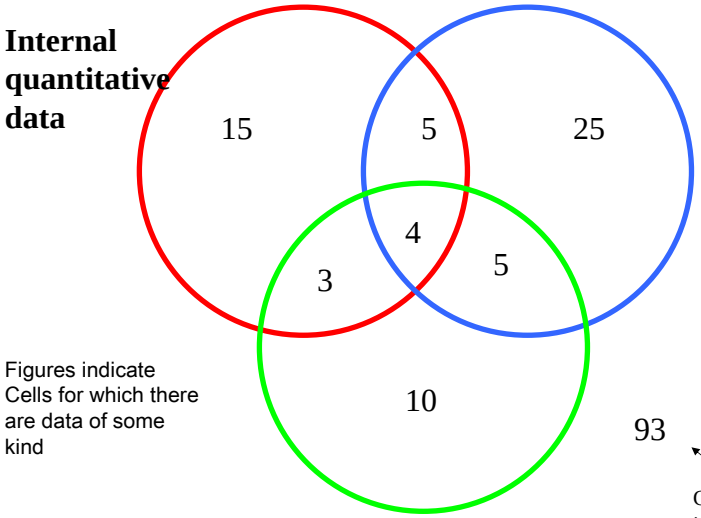
# Measurement units




Operational risk events are classified in *Business Lines* and *Event types*

For example we may consider 8 business lines and 20 event types, leading to 160 measurement units (named CELLS for brevity)

# Three sources of information

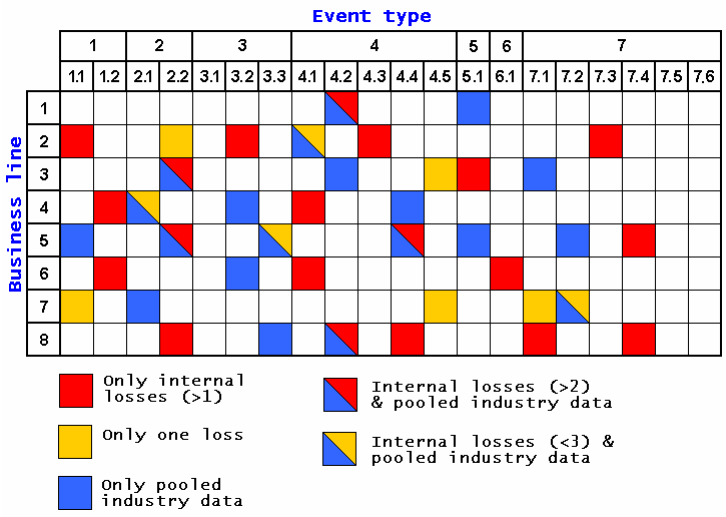



Figures indicate Cells for which there are data of some kind

Cells for which no information is available

23

# Non zero cells




24

# Internal Loss data: example

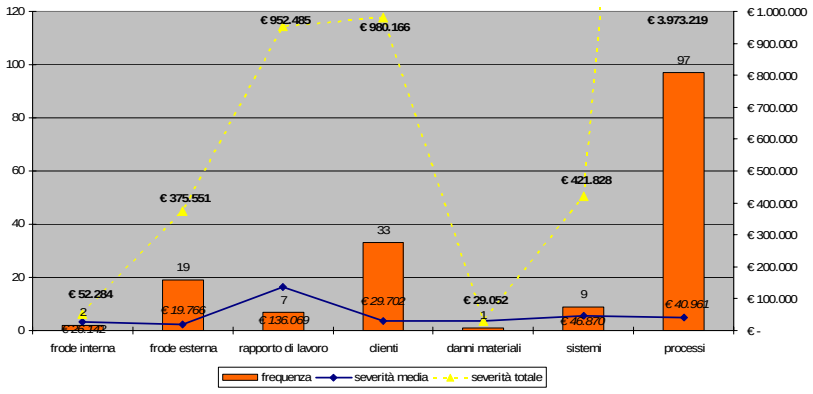
| | Total losses | | | | | Number of events (frequency) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Cells | 1999 | 2000 | 2001 | 2002 | 2003 | 1999 | 2000 | 2001 | 2002 | 2003 |
| Inc142 | 253,740 | 177,777 | 8,521 | 109,555 | 295,914 | 26 | 18 | 1 | 11 | 30 |
| Inc211 | 153,384 | 202,916 | 243,028 | 37,780 | 258,531 | 16 | 21 | 25 | 4 | 26 |
| Inc222 | 213,453 | 353,306 | 51,446 | 267,176 | 190,778 | 22 | 36 | 6 | 27 | 20 |
| Inc322 | 346,549 | 419,123 | 400,735 | 320,633 | 417,310 | 35 | 42 | 41 | 33 | 42 |
| Inc243 | 42,568 | 119,789 | 177,147 | 262,832 | 124,481 | 5 | 12 | 18 | 27 | 13 |
| Inc273 | 112,718 | 385,365 | 468,169 | 19,638 | 270,249 | 12 | 39 | 47 | 2 | 28 |
| Inc322 | 137,264 | 177,939 | 122,921 | 272,889 | 371,847 | 14 | 18 | 13 | 28 | 38 |
| Inc345 | 136,151 | 267,783 | 236,546 | 186,354 | 15,003 | 14 | 27 | 24 | 19 | 2 |
| Inc351 | 307,910 | 46,542 | 95,771 | 480,578 | 172,138 | 31 | 5 | 10 | 49 | 18 |
| Inc412 | 214,832 | 84,897 | 291,579 | 128,651 | 30,656 | 22 | 9 | 30 | 13 | 4 |
| Inc441 | 325,822 | 449,887 | 229,711 | 103,960 | 184,453 | 33 | 45 | 23 | 11 | 19 |
| Inc522 | 283,863 | 211,305 | 233,691 | 430,623 | 203,755 | 29 | 22 | 24 | 44 | 21 |
| Inc544 | 56,754 | 373,848 | 199,894 | 232,159 | 457,074 | 6 | 38 | 20 | 24 | 46 |
| Inc574 | 297,955 | 411,297 | 252,613 | 111,416 | 444,260 | 30 | 42 | 26 | 12 | 45 |
| Inc612 | 179,020 | 477,292 | 15,943 | 311,724 | 491,762 | 18 | 48 | 2 | 32 | 50 |
| Inc641 | 214,399 | 342,697 | 57,227 | 101,807 | 238,547 | 22 | 35 | 6 | 11 | 24 |
| Inc661 | 25,951 | 250,233 | 234,439 | 9,836 | 346,680 | 3 | 26 | 24 | 1 | 35 |
| Inc822 | 132,085 | 137,731 | 198,469 | 398,928 | 212,827 | 14 | 14 | | | |
| 942 | 199,389 | 114,443 | 416,757 | 137,136 | 391,998 | | | | | |
| | 418,052 | 7,217 | 380,512 | | | | | | | |

Not real data!

25

# External loss data: example

**Losses 2003-2006 for Italian Consortium DIPO**




Source: DIPO

26

# Self Assessment data: example

| Frequency | | 1.1.1 | 1.1.2 | 1.2.1 | 1.2.2 | 1.2.3 | 1.2.4 | 1.2.5 | 1.2.6 | 1.2.7 | 1.2.8 | 1.2.9 | 1.2.10 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | Annual | 30 | 30 | 28 | 32 | 27 | 30 | 32 | 32 | 29 | 32 | 27 | 29 |
| B | Monthly | 2 | 2 | 3 | | 3 | 2 | | | 3 | | 4 | 3 |
| C | Weekly | | | | | 2 | | | | | | 1 | |
| D | Daily | | | | | | | | | | | | |
| Total | | 32 | 32 | 31 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 |
| A | Annual | 93,8% | 93,8% | 90,3% | 100,0% | 84,4% | 93,8% | 100,0% | 100,0% | 90,6% | 100,0% | 84,4% | 90,6% |
| B | Monthly | 6,3% | 6,3% | 9,7% | 0,0% | 9,4% | 6,3% | 0,0% | 0,0% | 9,4% | 0,0% | 12,5% | 9,4% |
| C | Weekly | 0,0% | 0,0% | 0,0% | 0,0% | 6,3% | 0,0% | 0,0% | 0,0% | 0,0% | 0,0% | 3,1% | 0,0% |
| D | Daily | 0,0% | 0,0% | 0,0% | 0,0% | 0,0% | 0,0% | 0,0% | 0,0% | 0,0% | 0,0% | 0,0% | 0,0% |
| Total | | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |

27

# Risk indicators: example

$$F_{i,j} = \begin{cases} 1 + 0.3R_i^2 & \text{se } K_{i,j} > 0.66 \\ 1 & \text{se } 0.33 \leq K_{i,j} \leq 0.66 \\ 1 - 0.3R_i^2 & \text{se } K_{i,j} < 0.33 \end{cases}$$

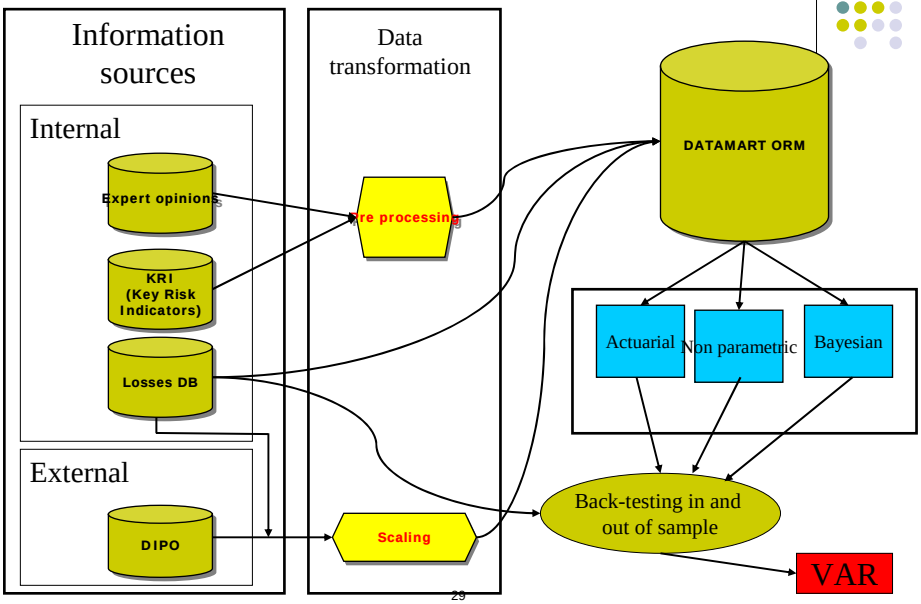
$F_{i,j}$ Correction factor of the i-th area (e.g. province) on the j-th cell (business line * event type)

$K_{i,j}$ Mean of the j-th KRI in the i-th area;

$R_i^2$ Ratio between deviance of the area means and total deviance for the i-th KRI;

28

# Structure of the proposed model



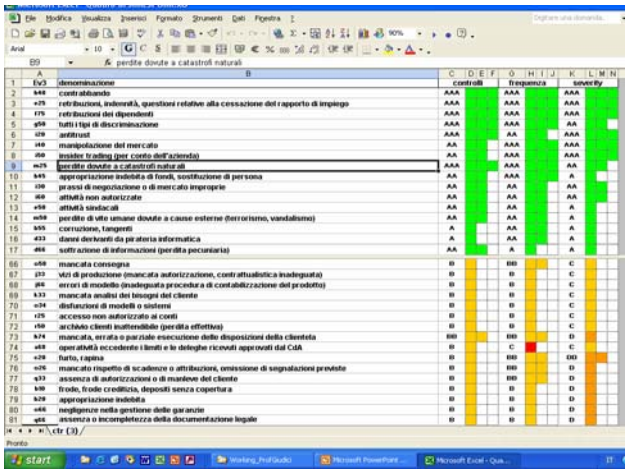

29

# Results – rating oprisks

| Frequency | | 1.1.1 | 1.1.2 | 1.2.1 | 1.2.2 | 1.2.3 | 1.2.4 | 1.2.5 | 1.2.6 | 1.2.7 | 1.2.8 | 1.2.9 | 1.2.10 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | Annual | 30 | 30 | 28 | 32 | 27 | 30 | 32 | 32 | 29 | 32 | 27 | 29 |
| B | Monthly | 2 | 2 | 3 | | 3 | 2 | | | 3 | | 4 | 3 |
| C | Weekly | | | | | 2 | | | | | | 1 | |
| D | Daily | | | | | | | | | | | | |
| **Total** | | **32** | **32** | **31** | **32** | **32** | **32** | **32** | **32** | **32** | **32** | **32** | **32** |
| A | Annual | 93,8% | 93,8% | 90,3% | 100,0% | 84,4% | 93,8% | 100,0% | 100,0% | 90,6% | 100,0% | 84,4% | 90,6% |
| B | Monthly | 6,3% | 6,3% | 9,7% | 0,0% | 9,4% | 6,3% | 0,0% | 0,0% | 9,4% | 0,0% | 12,5% | 9,4% |
| C | Weekly | 0,0% | 0,0% | 0,0% | 0,0% | 6,3% | 0,0% | 0,0% | 0,0% | 0,0% | 0,0% | 3,1% | 0,0% |
| D | Daily | 0,0% | 0,0% | 0,0% | 0,0% | 0,0% | 0,0% | 0,0% | 0,0% | 0,0% | 0,0% | 0,0% | 0,0% |
| **Total** | | **100%** | **100%** | **100%** | **100%** | **100%** | **100%** | **100%** | **100%** | **100%** | **100%** | **100%** | **100%** |
| Max | | 93,8% | 93,8% | 90,3% | 100,0% | 84,4% | 93,8% | 100,0% | 100,0% | 90,6% | 100,0% | 84,4% | 90,6% |
| Median | | A | A | A | A | A | A | A | A | A | A | A | A |
| Gini index | | **0,156** | **0,156** | **0,233** | **0,000** | **0,367** | **0,156** | **0,000** | **0,000** | **0,227** | **0,000** | **0,362** | **0,227** |
| Min Gini | | 0,000 | | | | | | | | | | | |
| Max Gini | | 0,849 | | | | | | | | | | | |
| **Gini Rating** | | **AAA** | **AAA** | **AAA** | **AAA** | **AA** | **AAA** | **AAA** | **AAA** | **AAA** | **AAA** | **AA** | **AAA** |

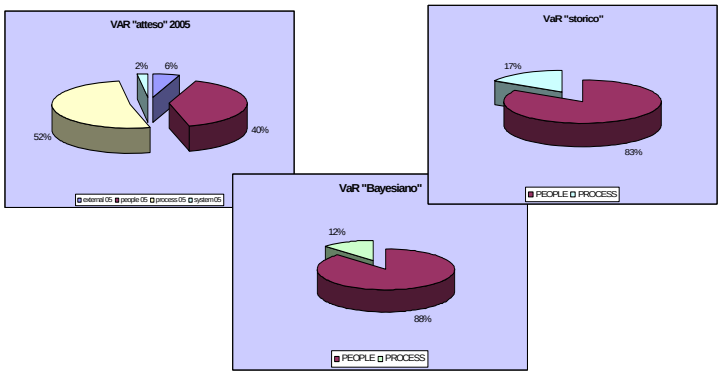
30

# Results - prioritisation



31

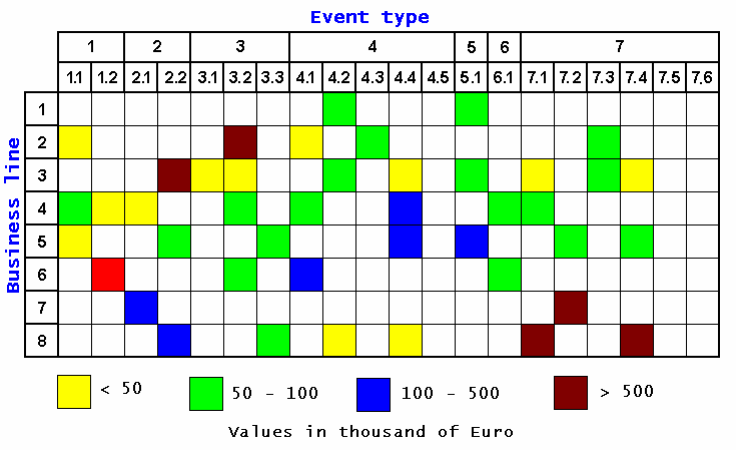
# Results – VaR reporting




- Comparison between expected, observed and non parametric VaR distributions

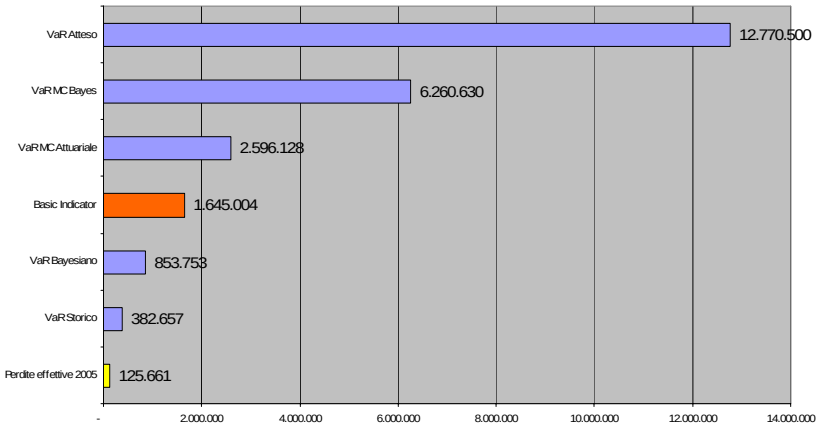
32

# Results - VaR reporting






# Results - Backtesting



# Merging data streams in operational risk management

## Open issues and references



## Open issues

a) Necessary to improve expert opinions: more reliable responses needed (cultural factor). For example, as the required VaR is at the 99,9% level, necessary to have different classes widhts (especially in the tails).

b) More work to calibrate different databases with each other (eg loss data collection, self-assessment, pooled data).

c) Necessary to measure and model explicitly key risk indicators and control factors, and insert them in the model

## References (1)


- Alexander, C. (2003). *Operational Risk: Regulation,Analysis and Management*, London. Financial Times, Prentice Hall.
- Basel Committee on Banking Supervision (2001) *Working Paper on the Regulatory Treatment of Operational Risk*, (www.bis.org).
- Bonafede, E.C. and Giudici, P. (2007). Bayesian networks for enterprise risk assessment. To appear in *Physica A.*
- Cornalba, C. and Giudici, P. (2004). Statistical models for operational risk management. *Physica A*, n. 338, pp. 166-172.
- Cruz M. (2002) *Modeling, Measuring and Hedging Operational Risk*, John Wiley & Sons, Chichester.
- Dalla Valle, L., Fantazzini, D. and Giudici, P. (2007). Copulae and Operational Risk. To appear in *International Journal of Risk Assessment and Management.*




## References (2)


- Fanoni, F., Giudici, P. and Muratori, M. (2005). Il rischio operativo: misurazione, monitoraggio, mitigazione. Il sole 24 ore.
- Frachot, A., Georges, P. and Roncalli, T. (2001). Loss distribution approach for operational risk. Working Paper, Groupe de Recherche Opérationnelle du Crédit Lyonnais.
- Giudici, P. and Bilotta, A. (2004). Modelling Operational Losses: a Bayesian Approach, *Quality and Reliability Engineering International*, n.20, pp. 407-417.
- Giudici P. (2003) *Applied Data Mining: Statistical Methods for Business and Industry*. John Wiley & Sons, Chichester.
- King J. (2001) *Operational Risk. Measurement and Modelling*, John Wiley & Sons, Chichester.
- Pezier, J. (2002). A constructive review of the Basel proposals on operational risk. ISMA Technical report, University of Reading.