AND OPERATIONAL RISK

Fabio Piacenza - UniCredit Operational Risk Methodologies and Control

First Milano R net meeting - May 8th 2012
INDEX

- Introduction
- AMA model
- Usage of R
- AMA model details with examples in R
UniCredit Group is a major international financial institution

- strong roots in 22 European countries
- more than 160,000 employees
- nearly 9,500 branches
- more than 35 million customers

Branches by country:

<table>
<thead>
<tr>
<th>Country</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Italy</td>
<td>4,400</td>
</tr>
<tr>
<td>Germany</td>
<td>1,030</td>
</tr>
<tr>
<td>Austria</td>
<td>310</td>
</tr>
<tr>
<td>Poland</td>
<td>953</td>
</tr>
<tr>
<td>Turkey</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>1,980</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>9,496</strong></td>
</tr>
</tbody>
</table>
The Risk Management function is mainly responsible for the following tasks:
- measurement and control of the institution exposure to the different types of risk;
- development and maintenance of systems for risk measurement, management and control, in agreement with the regulations.

On the basis of New Basel Capital Accord (Basel II) issued by the Basel Committee on Banking Supervision (an international body which coordinates operations of central banks and monetary authorities), financial institutions have to hold minimum capital requirements, separately for market risk, credit risk and operational risk.

**Strategic risk:** risk of losses from not doing the right thing

**Reputational risk:** risk of losses by not meeting shareholders’ expectations

**Credit risk:** risk of losses from borrowers not meeting their obligations
- It includes counterparty and country risk

**Operational risk:** risk of losses from inadequate or failed internal process, people and systems or from external events
- It includes legal risk but excludes strategic and reputational risks

**Market risk:** risk of losses from value changes of financial instruments
- It includes interest and exchange rate risks

**Other risks:** e.g. Business risk (risk of losses from business volume changes)
<table>
<thead>
<tr>
<th>OPERATIONAL RISK</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>People</strong></td>
</tr>
<tr>
<td>Fraud, collusion and other criminal activities</td>
</tr>
<tr>
<td>Violation of internal or external rules (unauthorized trading, insider dealing, etc.)</td>
</tr>
<tr>
<td>Breach of mandate</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
CAPITAL REQUIREMENT MEASUREMENT MODELS FOR OP. RISK

The international prudential rules require financial institutions to set aside a specific capital requirement for the exposure to Operational Risk choosing among three methods: Basic Indicator Approach - BIA, Standardized Approach - TSA, Advanced Measurement Approach - AMA

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
<th>Calculation</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIA</td>
<td>The capital requirement using the <strong>Basic Indicator Approach (BIA)</strong> is calculated by multiplying the three years average of relevant indicators* by a fixed percentage (15%).</td>
<td>RI(*) x 15%</td>
</tr>
<tr>
<td>TSA</td>
<td>The capital requirement using the <strong>Standardized Approach (TSA)</strong> is calculated by multiplying the three years average of the relevant indicator*, for each business line**, by a factor (12-15-18%) assigned to that business line.</td>
<td>RI(*) x 12-15-18%</td>
</tr>
<tr>
<td>AMA</td>
<td>With the <strong>Advanced Measurement Approach (AMA)</strong> the capital requirement is calculated with the bank’s internal operational risk measurement model, based on four elements (internal and external data, scenario analysis, business environment and internal control factors***).</td>
<td>Internal operational risk measurement model</td>
</tr>
</tbody>
</table>

(*) Relevant indicator is the gross income (item 120 of the Consolidated IAS Profit and Loss)
(**) Set of products or services offered by a company
(***) Usually implemented as “risk indicators”
INDEX

- Introduction

- AMA model

- Usage of R

- AMA model details with examples in R
AMA QUANTITATIVE REQUIREMENTS*

AMA models for the op. risk capital measurement shall be based on four elements

The internal loss data is a negative economic flow for which it is possible to identify the impact on the profit & loss account as consequence of an operational event.

A scenario is a fictitious operational event (also inspired from an occurred external event). The goal is to evaluate the impact in case the scenario occurs in the bank.

External data are operational loss events occurred to other financial institutions and banks.

Risk indicators are quantitative metrics reflecting operational risk exposure of specific processes or products.

Other quantitative requirements

- The approach must capture potentially severe tail loss events
- 1 year holding time period
- 99.9% confidence level
- Minimum 5 years observation period of internal loss data (3 years when the bank first moves to AMA)

(*) See the Circular 263 “New regulations for the prudential supervision of banks” issued by Bank of Italy (http://www.bancaditalia.it/vigilanza/normativa/norm_bi/circ-reg/vigprud/Circular_263_EN.pdf).
OPERATIONAL RISK CLASSES

An operational risk class is a data-homogeneous risk category, in terms of type, features or manifestation.

Examples of risk classes are:
- event type (ET);
- business line (BL);
- combination of business line/event type (BL x ET);
- products;
- legal entity class;
- etc.

Risk classes should be determined through analysis of homogeneity, sufficient data availability and independency.
AMA MODEL (OVERVIEW)

The commonly used approach to quantify Operational Risk is the Loss Distribution Approach, where frequency and severity of operational risk losses are modeled separately.

The yearly potential loss $S$ is based on the sum of the yearly losses $S_j$ related to the $J$ risk classes.

The yearly potential loss $S_j$, related to risk class $j$, is affected by two sources of uncertainty:

- The number of losses $N_j$ in one year time horizon
- The impact of each single loss $X_{ij}$

\[
S = \sum_{j=1}^{J} S_j
\]

\[
S_j = \sum_{i=1}^{N_j} X_{ij}
\]

(*) Risk indicators can be used to adjust the annual loss distribution of each risk class $j$, or directly the frequency and severity distributions.

(**) EL (expected loss) can be calculated as median, since the average could result too sensitive within extreme losses that cannot be considered as expected. Sufficient specific provisions should be available to deduct expected loss.
INDEX

- Introduction
- AMA model
- Usage of R
- AMA model details with examples in R
THE ROADMAP TO R IN UNICREDIT

2003: UniCredit started developing operational risk measurement model using S-Plus, a commercial software developed by Insightful Corporation. S-Plus is similar to R, since they are both derived from S language.

2007: UniCredit applied to Bank of Italy for the approval of AMA model with a calculation engine implemented in S-Plus

2008: UniCredit obtained the approval by Bank of Italy for using the AMA model to calculate the capital requirement for operational risk.

2008: TIBCO acquired Insightful Corporation

2009: UniCredit decided to migrate the calculation engine for operational risk from S-Plus to R

2010: UniCredit started the migration from S-Plus to R in cooperation with Quantide

2011: UniCredit, in cooperation with Quantide, developed a prototype in R of the core part (i.e. statistical functions) of the current calculation engine

2011: UniCredit started reviewing the AMA model

2011-2012: UniCredit is working on the reviewed AMA model using R to implement the calculation engine
**USAGE OF R IN UNICREDIT**

**Why R**

- Model building team in UniCredit was strongly experienced in S-Plus. After TIBCO acquired Insightful Corporation the progress development of S-Plus was basically arrested, because S-Plus was integrated in a more general software developed by TIBCO (Spotfire). The migration to R resulted to be the natural choice, since R is the language most similar to S-Plus (being both S-Plus and R derived from the S language)
- R is specific for statistical analysis
- R is open source
- R is supported and actively maintained by vast community of users
- Huge number of packages is available including most recent statistical techniques developed by research*
- A free and open source integrated development environment, named R-Studio (http://rstudio.org/), is available for R

**How R is used in UniCredit**

- Methodological research
- Development of prototypes
- Production environment
- Occasional data analysis

**Points of attention**

- It is difficult to follow the update software cycle, as we have to ensure reproducibility of all past results when we move to a new version of R
- Since R is an open source language, it is necessary to validate and benchmark the used packages (attention: several packages do the same things!)

(*) Available on CRAN package repository (http://cran.r-project.org/) .
INDEX

- Introduction
- AMA model
- Usage of R
- AMA model details with examples in R
LOSS SEVERITY

Since it can be difficult to model operational risk losses of a risk class using only one probability distribution, severity is usually analysed at two levels: body and tail of distribution (delimited by a high threshold value $u$).

The central body can be modelled by a parametric distribution (e.g. lognormal). Body is usually estimated on internal data, since their sample size is sufficient below the threshold $u$.

The tail can be modelled applying extreme value theory: the distribution that fits values above a high threshold is the Generalized Pareto (GPD). Tail is usually estimated on internal data integrated with external data and scenario generated data above $u$, since internal data are scarce above the threshold $u$. 
SEVERITY – BODY OF DISTRIBUTION

Usually, internal loss data are collected above a minimum threshold $H$.

To avoid biased estimates, maximum likelihood method conditioned to left and right truncations can be used:

$$f^*(x; \theta) = f(x; \theta | H \leq x \leq u) = \frac{f(x; \theta)}{Pr(x \leq u) - Pr(x \leq H)} = \frac{f(x; \theta)}{F(u; \theta) - F(H; \theta)} \quad H \leq x \leq u$$

where:
- $\theta$ is the vector including parameters of distribution (e.g. lognormal $\theta = (\mu, \sigma)$);
- $H$ is the minimum threshold (e.g. $H = \text{€}5,000$);
- $u$ is the threshold body-tail (e.g. $u = \text{€}1,500,000$);
- $f(x, \theta)$ is probability density function.

In case of lognormal:

$$f(x, \theta) = f_{LN}(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log(x) - \mu)^2}{2\sigma^2}\right)$$

The parameter estimate is obtained through maximization of the log-likelihood function:

$$l(\theta) = \sum_{i=1}^{n} \log\left(f^*(x_i, \theta)\right) \quad \Rightarrow \quad \hat{\theta} = \arg\max_{\theta} l(\theta)$$
SEVERITY – BODY OF DISTRIBUTION – EXAMPLE IN \( R \) (1/2)

```r
# Density function
ddist = function (x, distribution, theta1 = 0, theta2 = 0, L = 0, H = +Inf)
{
    ...
    if(distribution=="logNormal") out=dlnorm(x,theta1,theta2)/(plnorm(H,theta1,theta2)-plnorm(L,theta1,theta2))
    ...
    return(out)
}

# Neg-log-likelihood function
NlLike = function (theta, data, distribution, L = 0, H = +Inf)
{
    NlogL = -sum(log(ddist(x = data, distribution = distribution, theta1 = theta[1], theta2 = theta[2], L, H)))
    if (is.na(NlogL)) NlogL = .Machine$double.xmax
    return(NlogL)
}

# Max likelihood estimation
ldist = function(data, distribution, L = 0, H = +Inf, ...)
{
    ...
    if (distribution == "logNormal") {
        theta1 = mean(log(x[x > 0]))
        theta2 = sqrt(var(log(x[x > 0])))
    }
    ...
    theta = c(theta1, theta2)
    nlminbOut = optim(par = theta, fn = NlLike, method = "\...", control = ..., data = data, distribution = distribution, L = L, H = H)
    out = list(distribution = distribution, fit = nlminbOut)
}
```
# Example
set.seed(1000)
L=5
H=1500
x=rlnorm(n=10000, meanlog=2.5, sdlog=2)
x = x[x>L & x<H]
out=l dist(data=x,distribution="logNormal",L=L,H=H, outPlot = TRUE)
out

# Results
$\text{distribution}$
[1] "logNormal"

$\text{fit}$
$\text{fit}\$par
[1] 2.5005528 1.9948099

$\text{fit}\$value
[1] 33653.66

$\text{fit}\$counts
function gradient
7 2

$\text{fit}\$convergence
[1] 0

$\text{fit}\$message
NULL
SEVERITY – BODY OF DISTRIBUTION – SOME DISTRIBUTIONS

Lognormal (μ, σ) \[ f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{\log(x) - \mu}{2\sigma^2}\right) \]

Weibull (θ, τ) \[ f(x) = \frac{\tau(x/\theta)^{\tau - 1}e^{-(x/\theta)^\tau}}{x} \]

Loglogistic (γ, θ) \[ f(x) = \frac{\gamma(x/\theta)^\gamma}{x[1+(x/\theta)^\gamma]} \]

Example*

\[ \text{Lognormal}(2.501, 1.995) \]
\[ \text{Weibull}(0.338, 7.481) \]
\[ \text{Loglogistic}(0.846, 14.057) \]

(*) Plot of probability density functions of lognormal, Weibull and loglogistic distributions estimated with maximum likelihood method on data simulated from a lognormal(μ=2.5, σ=2)
SEVERITY – GOODNES-OF-FIT GRAPHICAL METHODS

Histogram and p.d.f. plot

Empirical vs theoretical c.d.f. plot

Q-Q plot

Comparison between theoretical and empirical probability density function.

Comparison between theoretical and empirical cumulative distribution function.

Comparison between theoretical and empirical quantiles. If points are near the bisector, fitting is good.

Q-Q plot provides the best evidence of the fitting on the tail region.
SEVERITY – GOODNESS-OF-FIT TESTS

The sorted sample \( x_1 \leq \ldots \leq x_n \) is supposed to be drawn from a population with distribution \( F(x) \)

**H0** (null hypothesis): sample follows distribution \( F(x) \)

**H1** (alternative hypothesis): sample does not follow distribution \( F(x) \)

<table>
<thead>
<tr>
<th>Kolmogorov – Smirnov test (KS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The test statistic ( KS^* ) is: ( D = \max_{1 \leq i \leq n} \left</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Anderson – Darling test (AD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The test statistic ( AD^{**} ) is: ( A^2 = n \int_{-\infty}^{\infty} \left[ F_n(x) - F(x) \right]^2 f[F(x)(1 - F(x))]dF(x) )</td>
</tr>
<tr>
<td>where ( F_n(x_i) = \frac{i}{n} ) is the empirical distribution function.</td>
</tr>
<tr>
<td>The statistic AD can be calculated as follows: ( A^2 = -n - \sum_{i=1}^{n} \frac{(2i - 1)}{n} \left[ \log F(x_i) + \log(1 - F(x_{n+1-i})) \right] )</td>
</tr>
</tbody>
</table>

**P-value** is calculated as the value such that \((1 - P-value)\) is maximum confidence level for the null hypothesis refusal.

Higher the **P-value**, better the goodness of fit given by the supposed distribution to the sample.

**If P-value \( \geq 0.05 \) there is not evidence for the null hypothesis refusal.**

(*) Maximum absolute difference between the empirical and the estimated distribution functions.

(**) Integral of the weighted squared difference between the empirical and the estimated distribution functions, where the weight is higher in the tail.
SEVERITY – GOODNESS-OF-FIT – EXAMPLE IN

# Example
library(opVar)  # package implemented by UniCredit in cooperation with Quantide*
set.seed(1000)
L = 5
H = 1500
x = rlnorm(n=10000, meanlog=2.5, sdlog=2)
xBody = x[x>L & x<H]
out = fittingResults(data=xBody, Distribution = c("logNormal", "weibull", "logLogistic"), L=L, H=H, createPlots=T)
out

# Results
> out
   Distribution Parameter1 Parameter2     KS.statistic KS.Pvalue AD.statistic AD.Pvalue Message
1       logNormal    2.5005528    1.994810      0.007899545   0.7976973   0.5286494 1.772738e-01 Converged
2      weibull       0.3376393    7.481478      0.008388092   0.7341320   0.5835832 1.288072e-01 Converged
3    logLogistic    0.8462676   14.066587      0.012133709   0.2790435   2.1340460 2.030057e-05 Converged

(*) Not available on CRAN package repository.
SEVERITY – TAIL OF DISTRIBUTION

On the basis of extreme value theory (EVT), the distribution function of loss data above a high threshold \( u \) is supposed to follow a GPD:

\[
G(x; \xi, \mu, \beta) = \begin{cases} 
1 - \left(1 + \xi \frac{x - \mu}{\beta}\right)^{-1/\xi} & \text{if } \xi \neq 0 \\
1 - e^{-(x-\mu)/\beta} & \text{if } \xi = 0
\end{cases}
\]

where:

\[
x \geq \mu \quad \text{if } \xi \geq 0
\]

\[
\mu \leq x \leq \mu - \frac{\beta}{\xi} \quad \text{if } \xi < 0
\]

The density function of GPD is:

\[
g(x; \xi, \mu, \beta) = \begin{cases} 
\frac{1}{\beta} \left(1 + \xi \frac{x - \mu}{\beta}\right)^{-1-1/\xi} & \text{if } \xi \neq 0 \\
\frac{1}{\beta} e^{-(x-\mu)/\beta} & \text{if } \xi = 0
\end{cases}
\]

where:

\[
x \geq \mu \quad \text{if } \xi \geq 0
\]

\[
\mu \leq x \leq \mu - \frac{\beta}{\xi} \quad \text{if } \xi < 0
\]

\( \xi \in \mathbb{R} \) and \( \beta > 0 \) are the respective shape and scale parameter.

\( \mu \) is the location parameter set equal to the threshold (\( \hat{\mu} = u \)).
SEVERITY – TAIL OF DISTRIBUTION – THRESHOLD SETTING

To estimate GPD parameters (i.e. \( \xi \) and \( \beta \)) we have to set a high threshold \( u \): the parameters are estimated on excess over threshold \( u \) using maximum likelihood estimation or other methods, e.g. probability weighted moments method (PWM), Hill estimator, etc.

**Mean excess function:**

\[
e(u) = E(X - u \mid X > u)
\]

For GPD:

\[
e(u) = \frac{\beta + \xi u}{1 - \xi}
\]

Since the mean excess function for GPD is a linear function of \( u \), if the empirical MEF is a straight line above a threshold, it’s an indication that the excess over \( u \) follows a GPD.
SEVERITY – TAIL OF DISTRIBUTION – PARAMETERS ESTIMATION

**Maximum likelihood method (ML)**

Considering the excess \( y_1 = x_1 - u, y_2 = x_2 - u, \ldots, y_n = x_n - u \) over the threshold \( u \), we can define the log-likelihood function to be maximized to obtain the parameters estimation:

\[
 l(\xi, \beta) = \sum_{i=1}^{n} \log(g_{\xi, \beta}(y_i)) \quad \Rightarrow \quad (\hat{\xi}, \hat{\beta}) = \arg \max_{\xi, \beta} l(\xi, \beta)
\]

**Probability weighted moments method (PWM)**

Considering the excess \( y_1 = x_1 - u, y_2 = x_2 - u, \ldots, y_n = x_n - u \) over the threshold \( u \) and the increasing-ordered excess \( y_{1:n}, y_{2:n}, \ldots, y_{n:n} \), parameter estimates are obtained as:

\[
 \hat{\xi} = 2 - \frac{\bar{y}}{\bar{y} - 2t} \quad \hat{\beta} = \frac{2 \cdot \bar{y} \cdot t}{\bar{y} - 2t}
\]

where:

\[
 \bar{y} = \sum_{i=1}^{n} y_i \quad t = \frac{\sum_{i=1}^{n} y_{i:n} \cdot (1 - p_{i:n})}{n} \quad p_{i:n} = \frac{i - 0.35}{n}
\]
# Example

library(evir)
set.seed(1000)
L=5
H=1500
x=rlnorm(n=10000, meanlog=2.5, sdlog=3)
nTail = length(x[x>H])
xBody = x[x>L & x<H]
xTail = rgpd(n=nTail, xi=0.5, mu=H, beta=1000)
x = c(xBody, xTail)
meplot(x, omit=length(x)/100)
abline(v = c(H), lty = 3)
title("Mean excess function")
parGpdMI = gpd(data=xTail, threshold=H, method="ml")$par.est$parGpdMI
parGpdPwm = gpd(data=xTail, threshold=H, method="pwm")$par.est
parGpdPwm
qplot(data=xTail,xi=parGpdMI[1],threshold=H)
title("Q-Q plot (ml)"
qplot(data=xTail,xi=parGpdPwm[1],threshold=H)
title("Q-Q plot (pwm")

# Results

> parGpdMI
  xi   beta
0.4878122 1025.8805247
> parGpdPwm
  xi   beta
0.4533816 1049.2300744
# Example
library(evir)
set.seed(1000)
L=5
H=1500
x=rlnorm(n=10000, meanlog=2.5, sdlog=2.5)
nTail = length(x[x>H])
xBody = x[x>L & x<H]
xTail = rgpd(n=nTail, xi=0.5, mu=H, beta=1000)
x = c(xBody, xTail)
meplot(x, omit=length(x)/100)
abline(v = c(H), lty = 3)
title("Mean excess function")
parGpdML = gpd(data=xTail, threshold=H, method="ml")$par.ests
parGpdML
parGpdPwm = gpd(data=xTail, threshold=H, method="pwm")$par.ests
parGpdPwm
qplot(data=xTail,xi=parGpdML[1],threshold=H)
title("Q-Q plot (ml)")
qplot(data=xTail,xi=parGpdPwm[1],threshold=H)
title("Q-Q plot (pwm)")

# Results
> parGpdML
  xi  beta
0.5542519  981.1148234
> parGpdPwm
  xi  beta
0.4911241 1025.0428461
Example: overwrite the qplot function of evir package

```r
qplot = function (data, xi = 0, trim = NA, threshold = NA, beta=1, line = TRUE, labels = TRUE, ...) {
  data = as.numeric(data)
  if (!is.na(threshold)) data = data[data >= threshold]
  if (!is.na(trim)) data = data[data < trim]
  if (xi == 0) {
    add = "Exponential Quantiles"
    y = qexp(ppoints(data))
  }
  if (xi != 0) {
    add = paste("GPD Quantiles; xi =", xi)
    y = qgpd(ppoints(data), xi = xi, mu=threshold, beta=beta)
  }
  plot(sort(data), y, xlab = "", ylab = "", ...) # Q-Q plot
  title("Ordered Data", ylab = add)
  if (line)
    abline(lsfit(sort(data), y))
    abline(0,1)
    invisible(list(x = sort(data), y = y))
}
qplot(data=xTail,xi=parGpdMl[1],threshold=H,beta=parGpdMl[2])
qplot(data=xTail,xi=parGpdPwm[1],threshold=H,beta=parGpdPwm[2])
```

Example (1/3) LN(2.5,3)

Example (2/3) LN(2.5,2.5)
LOSS FREQUENCY

A standard choice to estimate annual frequency of operational risk loss events is the Poisson distribution*

A discrete random variable $N$ follows a Poisson distribution with parameter $\lambda$, $N \sim \text{Po}(\lambda)$, if it has the following probability mass function:

$$\Pr(N = n) = e^{-\lambda} \frac{\lambda^n}{n!} \quad n = 0,1,2,...$$

Given the sample $n_1, \ldots, n_t$ drawn from a distribution $\text{Po}(\lambda)$, maximum likelihood estimate of parameter $\lambda$ is given by the empirical mean:

$$\hat{\lambda}_{\text{sample}} = \frac{1}{t} \sum_{i=1}^{t} n_i$$

The parameter can be estimated on internal loss data higher or equal to the minimum threshold $H$

Parameter estimate has to be adjusted for reporting bias:

$$\hat{\lambda} = \frac{\hat{\lambda}_{\text{sample}}}{\Pr(x > H)} = \frac{\hat{\lambda}_{\text{sample}}}{1 - F(H)}$$

where $F(H)$ is the c.d.f. of severity distribution calculated in the minimum threshold $H$

(*) An alternative is the negative binomial distribution. However, since the choice of frequency distribution analytical form has not a significant impact on the capital requirement, Poisson distribution is usually preferred because it is simpler, having only one parameter (representing both mean and variance), instead of the two parameters of negative binomial distribution.
# Example

```r
set.seed(1000)
L = 5 # minimum loss threshold
ni = c(550,700,380,430,600) # number of losses per year above minimum threshold
lambdaSample = mean(ni) # parameter of Poisson distribution based on losses above minimum threshold
lambda = lambdaSample / (1 - pnorm(q=L, meanlog=2.5, sdlog=2)) # parameter of Poisson adjusted for reporting bias

hist(x=rpois(n=10000,lambda=lambda), 50, xlab="N events / Year", ylab="Probability", prob=TRUE, main="")
title("Frequency distribution")

lambdaSample
lambda
```

# Results

```r
> lambdaSample
[1] 532
> lambda
[1] 791.7354
```
CONVOLUTION: MONTE CARLO METHOD

1. Extract one random number \( n_j \) from frequency distribution, e.g. \( n_j = 3 \)

2. Extract \( n_j \) random numbers \( (x_{ij}) \) from severity distribution, e.g. 
\[
x_{1j} = 12000 \quad x_{2j} = 18000 \quad x_{3j} = 10000
\]

3. Sum: obtain a possible figure of yearly op. loss \( s_j = \sum x_{ij} \)
\[
e.g. \quad s_j = x_{1j} + x_{2j} + x_{3j} = 40000
\]

Repeat e.g. 5 mln times
CONVOLUTION: MONTE CARLO METHOD – EXAMPLE IN

# Example

library(evir)

# Quantile function of lognormal-GPD severity distribution
qlnorm.gpd = function(p, theta, theta.gpd, u){
    Fu = plnorm(u, meanlog=theta[1], sdlog=theta[2])
    x = ifelse(p<Fu,
        qlnorm( p=p, meanlog=theta[1], sdlog=theta[2] ),
        qgpd( p=(p - Fu) / (1 - Fu) , xi=theta.gpd[1], mu=theta.gpd[2], beta=theta.gpd[3]) )
    return(x)
}

# Random sampling function of lognormal-GPD severity distribution
rlnorm.gpd = function(n, theta, theta.gpd, u){
    r = qlnorm.gpd(runif(n), theta, theta.gpd, u)
}

set.seed(1000)
nSim = 1000000 # Number of simulated annual losses
H = 1500 # Threshold body-tail
lambda = 791.7354 # Parameter of Poisson body
theta1 = 2.5 # Parameter mu of lognormal (body)
theta2 = 2 # Parameter sigma of lognormal (body)
theta1.tail = 0.5 # Shape parameter of GPD (tail)
theta2.tail = H # Location parameter of GPD (tail)
theta3.tail = 1000 # Scale parameter of GPD (tail)
sj = rep(0,nSim) # Annual loss distribution inizialization
freq = rpois(nSim, lambda) # Random sampling from Poisson
for(i in 1:nSim) # Convolution with Monte Carlo method
    sj[i] = sum(rlnorm.gpd(n=freq[i], theta=c(theta1,theta2), theta.gpd=c(theta1.tail, theta2.tail, theta3.tail), u=H))

Annual loss distribution
OVERALL LOSS DISTRIBUTION BASED ON COPULA FUNCTION

Overall loss distribution, starting from loss distributions related to risk classes, can be obtained through a copula function such as Gaussian, Student-t or Gumbel copulas.

Copula is a multivariate distribution function, $C$, with marginal functions distributed uniformly in $[0,1]$ ($U(0,1)$) such that:

1. $C : [0,1]^n \rightarrow [0,1]$
2. $C$ has marginal functions $C_i$ such that $C_i(u) = C(1,...,1,u,1,...,1) = u \quad \forall u \in [0,1]$

**Sklar’s theorem:**

$$F(x_1,\ldots,x_n) = C(F_1(x_1),\ldots,F_n(x_n))$$

where $F$ is a $n$-dimensional distribution function with marginal functions $F_1,\ldots,F_n$

---

**Corollary of Sklar’s theorem:**

$$C(u_1,\ldots,u_n) = F(F_1^{-1}(u_1),\ldots,F_n^{-1}(u_n)) \quad \forall u = (u_1,\ldots,u_n) \in [0,1]^n$$
OVERALL LOSS DISTRIBUTION: STUDENT-T COPULA

Student-t copula is the copula function of multivariate Student t distribution.

Let \( X = (X_1, \ldots, X_n) \) be a vector with a multivariate t distribution with \( \nu \) degrees of freedom. The copula of vector \( X \) can be analytically represented in the following form:

\[
C^t_{\nu, R}(u_1, \ldots, u_n) = t_{\nu, R}(t^{-1}_{\nu}(u_1), \ldots, t^{-1}_{\nu}(u_n))
\]

where:
- \( R \) is the correlation matrix*;
- \( \nu \) represents the number of degrees of freedom**;
- \( t^n_{\nu, R} \) denotes the multivariate t distribution function;
- \( t^1_{\nu} \) denotes the margins of \( t^n_{\nu, R} \).

**Examples in 2-dimensions (i.e. 2 risk classes case)**

\[
\begin{array}{ccc}
\text{R}_{12}=0.8 & \nu =10 \\
\text{R}_{12}=0.8 & \nu =3 \\
\text{R}_{12}=0.5 & \nu =10 \\
\text{R}_{12}=0.5 & \nu =3
\end{array}
\]

(*) Correlation matrix can be estimated through a transformation of Kendall (or Spearman) rank correlation among the observed yearly aggregated losses of the different risk classes (monthly aggregated losses could be used to increase data availability, in case they are not significantly autocorrelated).

(**) Degrees of freedom can be estimated with maximum likelihood method.
OVERALL LOSS DISTRIBUTION: MONTE CARLO METHOD

1. Extract one random value \( \overline{u} = (u_1, \ldots, u_j) \) from copula function, e.g., supposing \( J = 2 \), \( u_1 = 0.6 \), \( u_2 = 0.9 \).

2. Extract \( J \) random numbers \( s_j = F_{S_j}^{-1}(u_j) \) \((j = 1, \ldots, J)\) from loss distributions related to the \( J \) risk classes, e.g.,
   \[
   s_1 = F_{S_1}^{-1}(0.6) = 300000 \quad s_2 = F_{S_2}^{-1}(0.9) = 500000
   \]

3. Sum: obtain a possible figure of overall yearly op. loss \( s = \sum_j s_j \)
e.g. \( s = s_1 + s_2 = 300000 + 500000 = 800000 \)

(*) The plot can be drawn only in 2-dimensions case (i.e. 2 risk classes case).
# Example
library(QRM)
set.seed(1000)
nSim = 1000000 # Number of simulated overall annual losses
s1 = rlnorm(n=nSim, meanlog=4.5, sdlog=2.3) # Loss distribution risk class 1
s2 = rlnorm(n=nSim, meanlog=5, sdlog=2.5) # Loss distribution risk class 2
VaR.s1 = quantile(s1, 0.999) # VaR risk class 1
VaR.s2 = quantile(s2, 0.999) # VaR risk class 2
corr = 0.6 # Correlation among risk classes
corrMatrix = matrix(data=c(1,corr,corr,1), nrow=2) # correlation matrix
dof = 5 # degrees of freedom
simCopulaT = rcopula.t(n=nSim, df=dof, Sigma=corrMatrix) # Simulations from Student-t copula
s = quantile(s1, simCopulaT[,1]) + quantile(s2, simCopulaT[,2]) # overall annual loss distribution
VaR.s = quantile(s, 0.999)
divEff = (VaR.s1+VaR.s-VaR.s)/(VaR.s1+VaR.s) # diversification effect
EL = quantile(s, 0.5) # Expected loss
capReq = VaR.s - EL # Capital requirement
VaR.s1; VaR.s2; VaR.s; divEff; EL; capReq

# Results
> VaR.s1; VaR.s2; VaR.s; divEff; EL; capReq
111300.8
339577.2
410748.6
0.2131997
355.3662
410393.2
### SOME USEFUL PACKAGES FOR OPERATIONAL RISK

<table>
<thead>
<tr>
<th>Package name</th>
<th>Scope</th>
<th>Usage in operational risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>fitdistrplus</td>
<td>Help to fit of a parametric distribution to non-censored or censored data</td>
<td>Several functions to fit parametric distributions to non-censored or censored data.</td>
</tr>
<tr>
<td>VGAM</td>
<td>Vector Generalized Linear and Additive Models</td>
<td>Density, distribution function, quantile function and random generation for several distributions</td>
</tr>
<tr>
<td>truncgof</td>
<td>GoF tests allowing for left truncated data</td>
<td>Check goodness of fit related to left truncated distributions</td>
</tr>
<tr>
<td>POT</td>
<td>Generalized Pareto Distribution and Peaks Over Threshold</td>
<td>Estimate tail of severity distribution</td>
</tr>
<tr>
<td>evir</td>
<td>Extreme Values in R</td>
<td>Estimate tail of severity distribution</td>
</tr>
<tr>
<td>fExtremes</td>
<td>Rmetrics - Extreme Financial Market Data</td>
<td>Estimate tail of severity distribution</td>
</tr>
<tr>
<td>extRRemes</td>
<td>Extreme value toolkit</td>
<td>Estimate tail of severity distribution</td>
</tr>
<tr>
<td>copula</td>
<td>Multivariate Dependence with Copulas</td>
<td>Copula functions</td>
</tr>
<tr>
<td>QRM</td>
<td>Provides R-language code to examine Quantitative Risk Management concepts</td>
<td>Estimate tail of severity distribution and copula functions</td>
</tr>
<tr>
<td>ROracle / RODBC / RJDBC</td>
<td>OCI based Oracle database interface for R / ODBC Database Access / Provides access to databases through the JDBC interface</td>
<td>Read operational risk data from databases</td>
</tr>
</tbody>
</table>
Many thanks for your attention

Fabio.Piacenza@unicredit.eu