

HSBC Quants Academy

Operational Risk Modelling

Prepared by: Katarzyna Potaśnik

Date: 18 May 2026



Disclaimer

This presentation has been prepared by HSBC Service Delivery (Polska) Sp. z o.o. with a registered office located at Kapelanka 42A, 30-347 Kraków, Poland, registered in the District Court for Kraków-Śródmieście in Kraków, XI Commercial Division of the National Court Register, under the number KRS 0000310459, share capital (entirely paid) in the amount of 67 689 500,00 PLN, NIP 525-242-91-07, hereinafter referred to as "HSBC".

HSBC has based this document on information obtained from sources it believes to be reliable but which has not been independently verified. While this information has been prepared in good faith, no representation or warranty, express or implied, is or will be made and no responsibility or liability is or will be accepted by the HSBC, or by any of their respective officers, employees or agents in relation

to the accuracy or completeness of the Documents or any other written or oral information made available to any interested party or its advisers and any such liability is expressly disclaimed. Any charts and graphs included are from publicly available sources or proprietary data. No liability is accepted whatsoever for any direct, indirect or consequential loss arising from the use of this document. HSBC is under

no obligation to keep current the information in this document. You are solely responsible for making your own independent appraisal of

and investigations into the data, products, financial instruments and transactions referred to in this document and you should not rely on

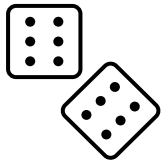
any information in this document as constituting investment advice. Neither HSBC nor any of its affiliates are responsible for providing you with legal, tax or other specialist advice and you should make your own arrangements in respect of this accordingly.

The issuance of and details contained in this document, which is not for public circulation, does not constitute an offer or solicitation for, or

advice that you should enter into, the purchase or sale of any security, commodity or other financial instrument or master agreement, or any other contract, agreement or structure whatsoever. This document is intended to be distributed in its entirety. Reproduction of this document, in whole or in part, or disclosure of any of its contents, without prior consent of HSBC or any associate, is prohibited.

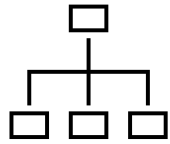
Agenda

**Operational Risk
Loss Definition**



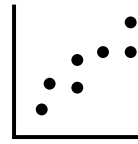
1

**Model Purpose, Scope
and Design**



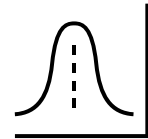
2

Data Profile



3

**Modelling Extreme
Events**



4

Operational Risk Loss

Definition of operational risk

Operational risk is the risk of losses caused by flawed or failed processes, policies, systems or events that disrupt business operations. Employee errors, criminal activity such as fraud, and psychical events are among the factors that can trigger operational risk.

Basel definition of operational risk

Operational risk is defined as the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events. This definition includes legal risk but excludes strategic and reputational risk.

Root cause of operational loss:

- ❖ New laws or regulatory requirements
- ❖ Regulatory compliance violations
- ❖ Failure to adhere to the company's policies
- ❖ Human errors
- ❖ Man-made disasters, such as terrorism or cybercrime
- ❖ Natural disasters
- ❖ Many others

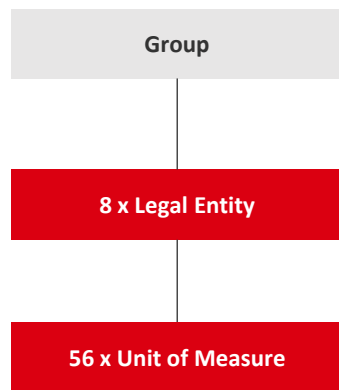
Basel Risk Categories

- ❖ Internal Fraud
- ❖ External Fraud
- ❖ Employment Practices and Workplace Safety
- ❖ Clients, Products and Business Practices
- ❖ Damage to Physical Assets
- ❖ Business Disruption and System Failures
- ❖ Execution, Delivery, and Process Management

Operational Risk Economic Capital Model

Structure and Purpose

Model granularity



Legal Entities (LE) identify regions.

Unit of Measure (UoM):
LE x Risk Category

Purpose

- ◆ The model aims to provide conservative estimates of the operational risk economic capital requirements over a one-year time horizon. Economic capital is estimated to cover losses in the next year with at least 99.9% likelihood (i.e. less than 0.1% likelihood of higher losses), less expected losses that are budgeted for.
- ◆ The material unit of the model is the 99.9th VaR of the aggregate loss distribution for each Legal Entity and for the Group as a whole. The model aims to capture tails events and respond to the major drivers of risk affecting the tail distribution.

Assumption about annual loss distribution

- ◆ Each Unit of Measure is modelled as a compound distribution of the number of losses (frequency distribution) and the size of each individual loss (severity distribution).

$$\text{Annual Loss} = L_1 + L_2 + \dots + L_N$$

where the losses L_i are modelled using a severity distribution and frequency N follows Poisson distribution. N and L_i where $i \in \mathbb{N}$ are assumed to be independent.

- ◆ The annual loss distribution is therefore a convolution of the same severity distribution, and it is simulated using Monte Carlo methods.

Operational Risk Economic Capital Model

Annual Loss Distribution

The statistical model simulates potential losses in a year.

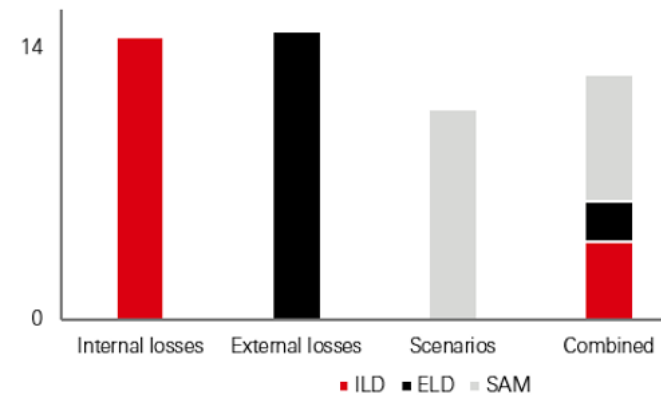
- ◆ Simulates a random number of events in a year and the severity of each event
- ◆ Economic Capital is set to cover 99.9% of the potential losses (i.e. all but the worst 0.1%), less the expected loss



The model estimates potential losses using three alternative methods then weights between them:

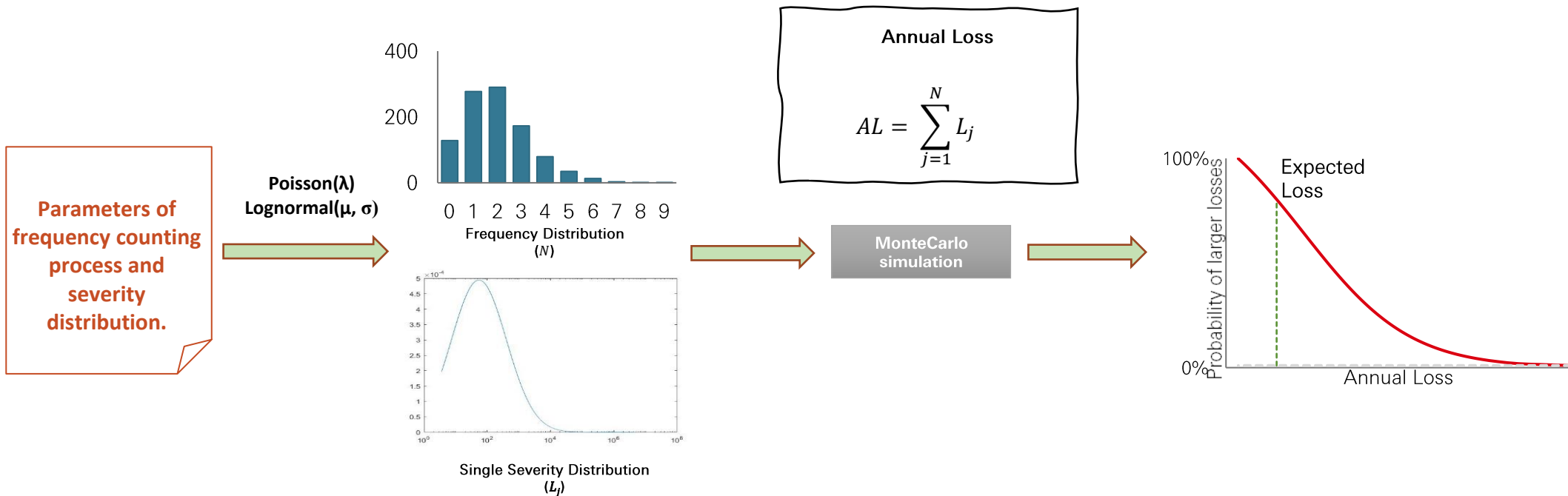
- ◆ Internal Loss Data : w1
- ◆ External Loss Data : w2
- ◆ Scenario Approach Model : w3

Further capital is added for risks and entities not covered by the model



Operational Risk Economic Capital Model

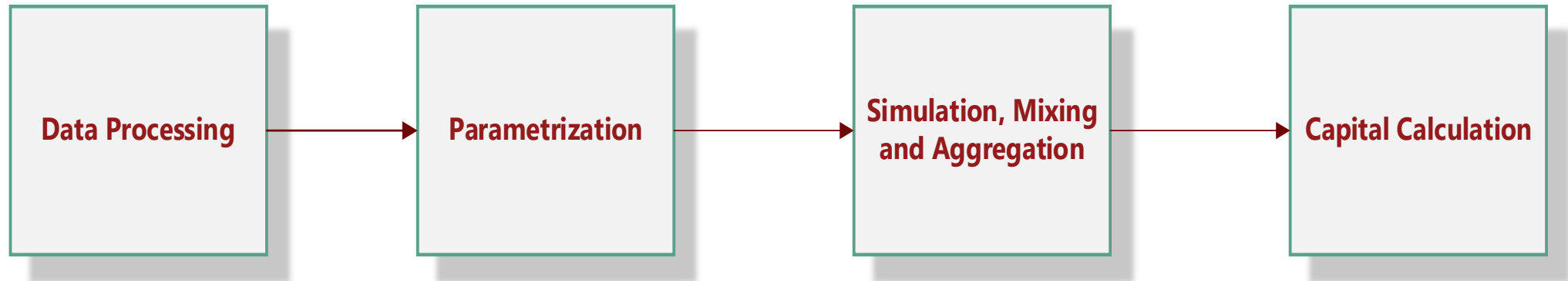
Compound Loss Distribution



Operational Risk Economic Capital Model

Model Design

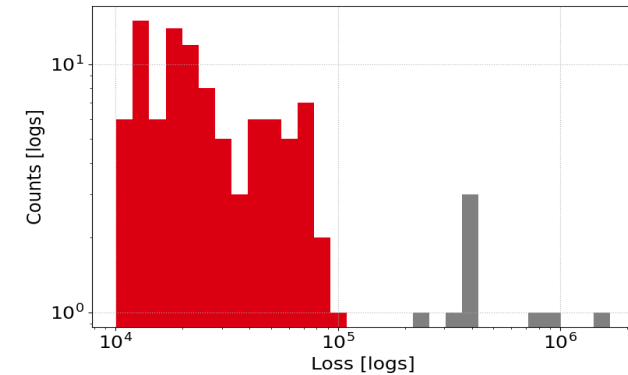
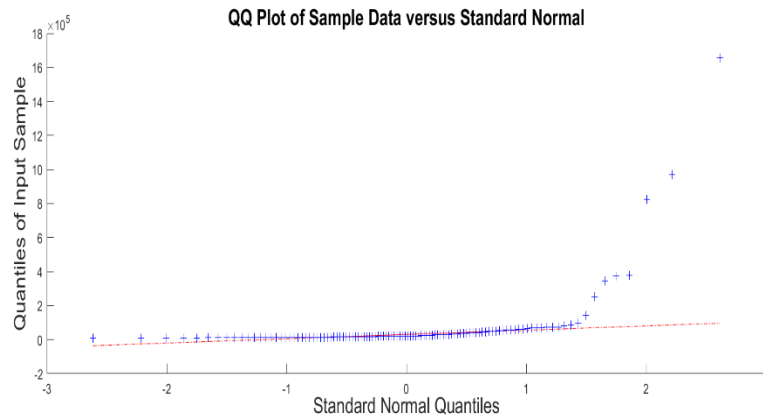
Building blocks of the model



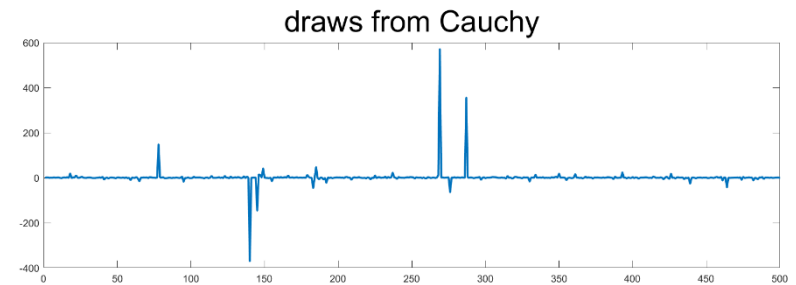
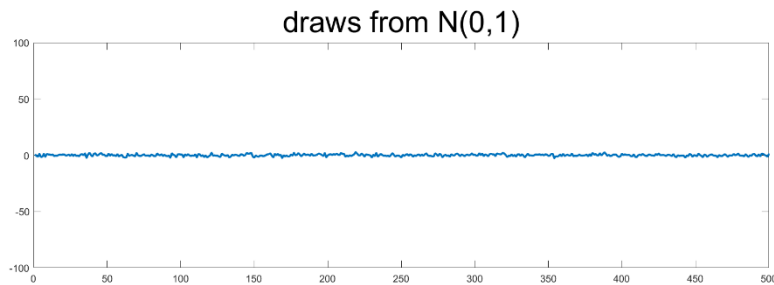
Operational Risk Data

Typical operational risk dataset contains:

- ◆ High frequency low severity losses
- ◆ Low frequency high severity losses



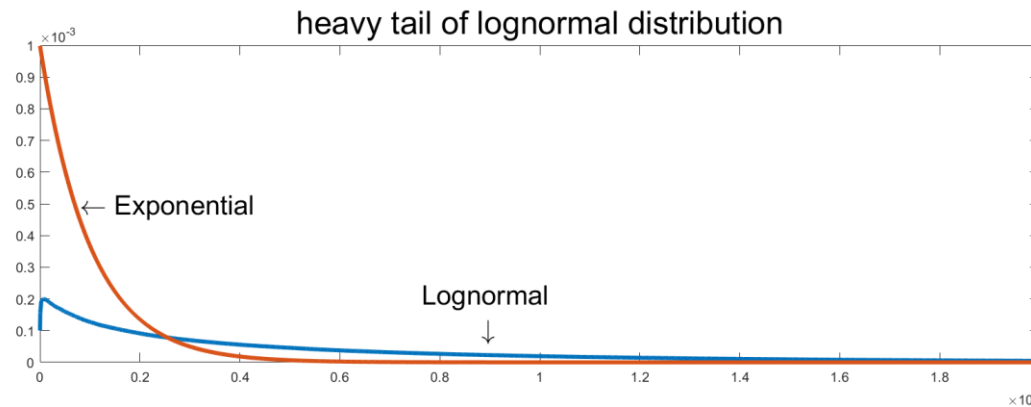
Another example of heavy-tailed distributed data



Modelling extreme events

Heavy-tailed distributions

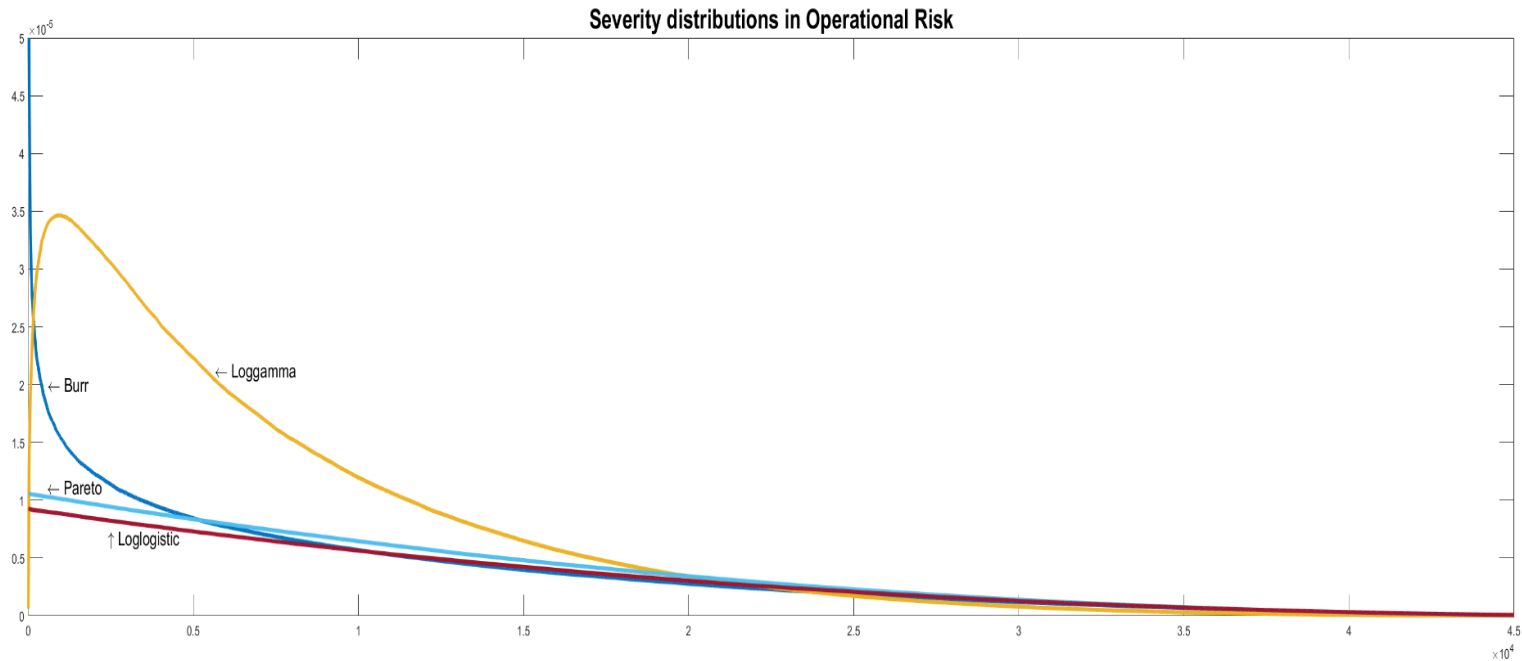
- ◆ A heavy-tailed distribution has tails that are heavier than an exponential distribution (tail part decays more slowly than the exponential distribution)



- ◆ Heavy-tailed distributions are used for modelling processes that may produce extreme outcomes
- ◆ Fat-tailed distributions is a subclass of the heavy-tailed distributions;
The distribution of X has a fat right tail if there is a positive exponent $\alpha > 0$ called the tail index such that $P(X > x) \sim x^{-\alpha}$
- ◆ In Operational Risk we use distributions with fat right tails

Modelling extreme events

Severity distribution families



Single distributions

- ◆ Lognormal
- ◆ Weibull
- ◆ Generalized Pareto
- ◆ Burr
- ◆ Loggamma
- ◆ Loglogistic
- ◆ Pareto

Spliced distributions

- ◆ Lognormal- Generalized Pareto

Modelling extreme events

Parametrization Process

Fitting the severity distribution

Fitting severity distribution to data is performed with the Maximum-Likelihood Estimator (MLE). It maximizes the PDF for the observed data, i.e. for given losses $(X_1, \dots, X_n)$ that are assumed to be a random sample from $f_X(\cdot; \theta)$, $\theta \in \Theta$, the function $L(\theta) = \prod_{i=1}^n f(x_i|\theta)$ is maximized for $\theta \in \Theta$. This is achieved by solving the equation

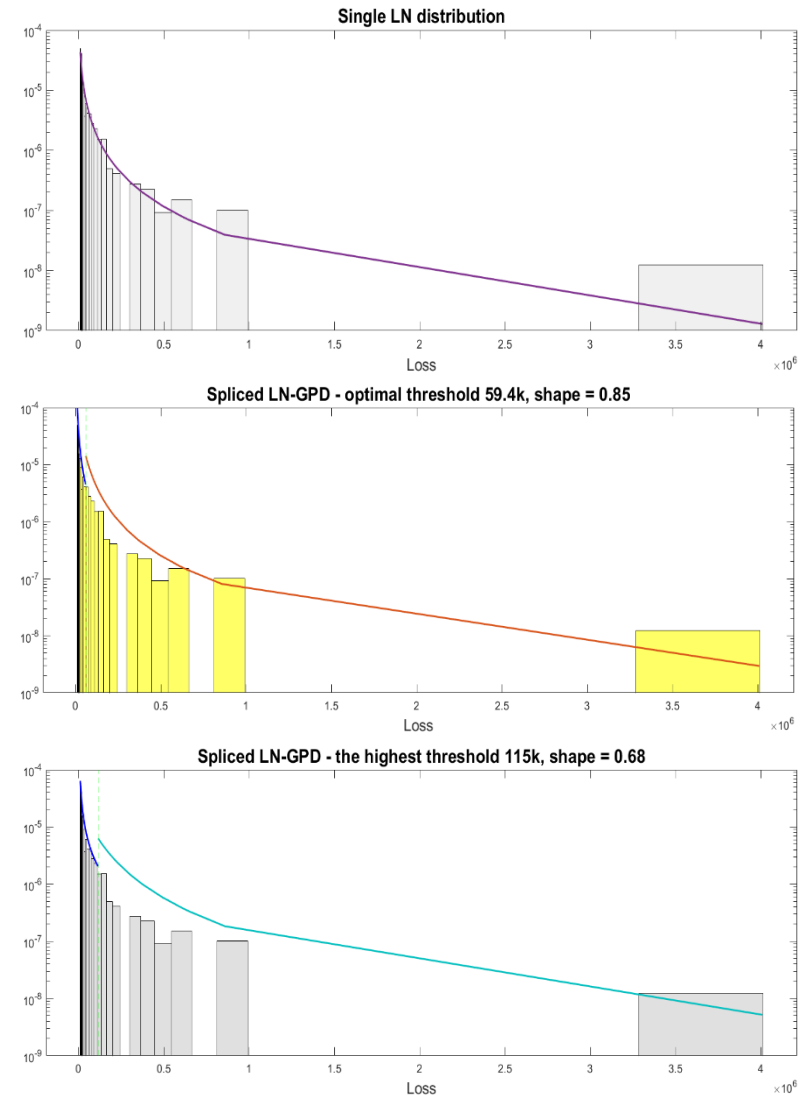
$$\frac{\partial \text{Log}L(\theta)}{\partial \theta} = \sum_{i=1}^n \frac{\partial}{\partial \theta} \log(f(x_i|\theta)) = 0, \quad \theta \in \Theta.$$

Challenges of the fitting process:

- If a single distribution does not fit the data, a spliced distribution is used with separate distributions for the body and the tail:

$$f(x; \theta_B, \theta_T) = \begin{cases} f_B(x; \theta_B), & 0 \leq x \leq T \\ f_T(x; \theta_T), & T < x < +\infty \end{cases}$$

- Fitting distributions with infinite variance
- Searching for optimal threshold when splicing a distribution
- Selection criteria for the optimal threshold: metric based on KS, CvM, AD, AD2 goodness of fit tests.



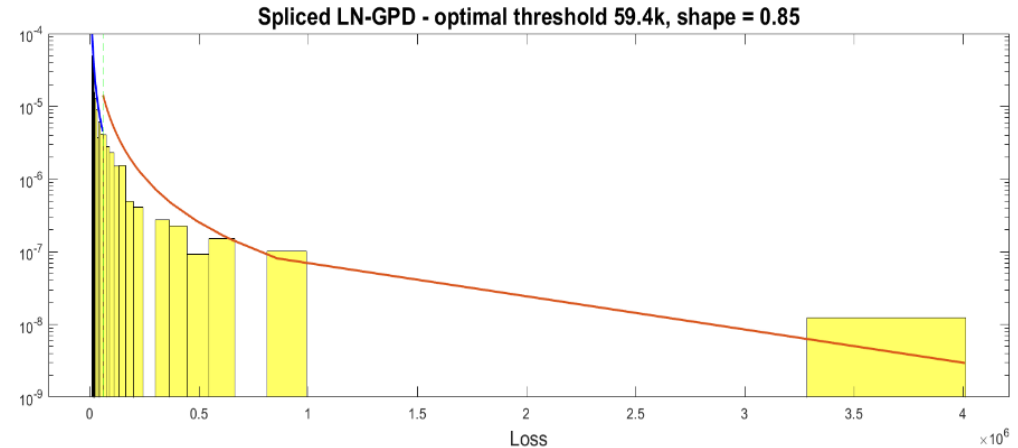
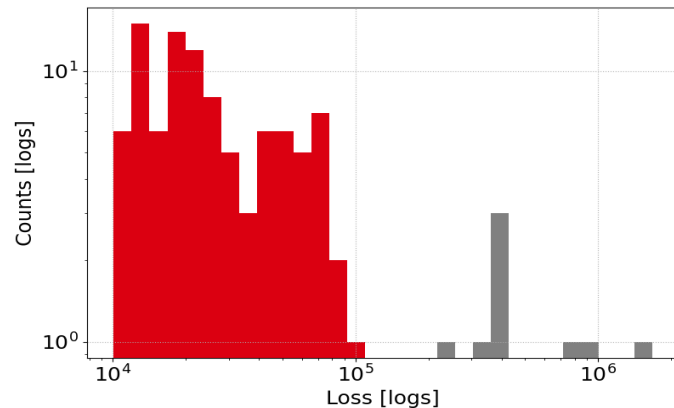
Modelling extreme events

Spliced severity distribution

Spliced distributions combine two or more different probability distributions at a predefined splicing points. This allows model to represent data that has distinct behaviour in different ranges like for example heavier tail in one region.

In Operational Risk EC Model spliced distributions combine Lognormal as a „body distribution“ f_B with Generalized-Pareto as a „tail distribution“ f_T in a body-tail threshold T .

$$f(x; \theta_B, \theta_T) = \begin{cases} f_B(x; \theta_B), & 0 \leq x \leq T \\ f_T(x; \theta_T), & T < x < +\infty \end{cases}$$



Modelling extreme events

Non - Parametric Outlier Detector

1

Hypothesis

H_0 : there is no outlier in a sample
 H_1 : there is at least one outlier in a sample

2

Test Statistics

$$\hat{D}_{J_n} = \log(2) \frac{\max_{j \in \{1, \dots, J_n\}} (j \cdot \log(\tau_{(n-j)}))}{\text{median}_{j \in \{1, \dots, J_n\}} (j \cdot \log(\tau_{(n-j)}))}, \text{ where:}$$

- $\tau_{(n-j)}$ is a $n - j$ ratio
- J_n is function of sample size.
- $P(\hat{D}_{J_n} \leq x) \sim (1 - e^{-x})^{J_n}, n \rightarrow \infty$

3

Decision

Critical value: $C_\alpha = -\log(1 - (1 - \alpha)^{1/J_n})$

- if $\hat{D}_{J_n} > C_\alpha$ hypothesis H_0 is rejected,
- the largest $\hat{k}_0$ such that:

$$\frac{\hat{k}_0 \cdot \log(\tau_{(n-\hat{k}_0)})}{\text{median}_{j \in \{1, \dots, J\}} (j \cdot \log(\tau_{(n-j)}))} > C_\alpha$$

estimates the number of outliers.

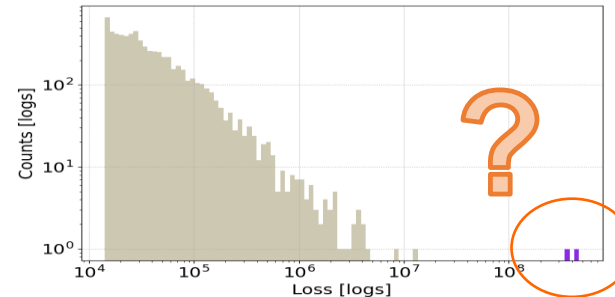
- Given a vector of sorted observations $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$, we would like to detect whenever there is a gap between $X_{(i)}$ drawn from a common distribution and one or few $X_{(j)}$ generated by more severe distributions,
- The test is built on ratios $\tau_{(n-j)} = X_{(n-j+1)} / X_{(n-j)}, j = 1, \dots, n - 1$ from highest loss downwards,
- Asymptotic distribution of the test statistics $\hat{D}_{J_n}$ follows $P(\hat{D}_{J_n} \leq x) \sim (1 - e^{-x})^{J_n}, n \rightarrow \infty$,
- To make a test decision, we compare test statistics to critical value C_α ,
- If H_0 is rejected then the largest index $\hat{k}_0$ such that $\frac{\hat{k}_0 \cdot \log(\tau_{(n-\hat{k}_0)})}{\text{median}_{j \in \{1, \dots, J\}} (j \cdot \log(\tau_{(n-j)}))} > C_\alpha$ indicates the number of detected outliers.

Modelling extreme events

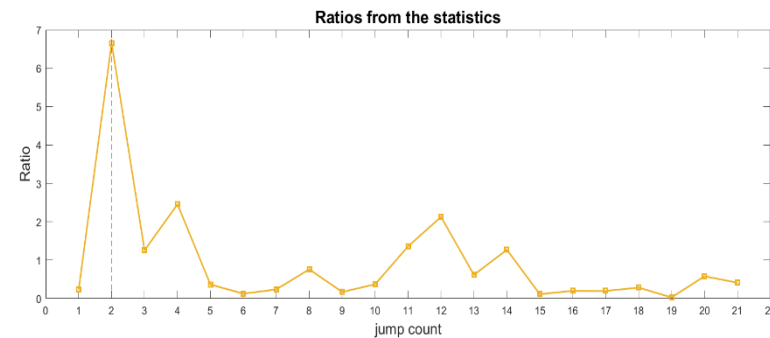
Non - Parametric Outlier Detector

Non Parametric Outlier Detector (NPOD) not only allows a decision of the hypothesis test problem, but also for identification of the exact outliers.

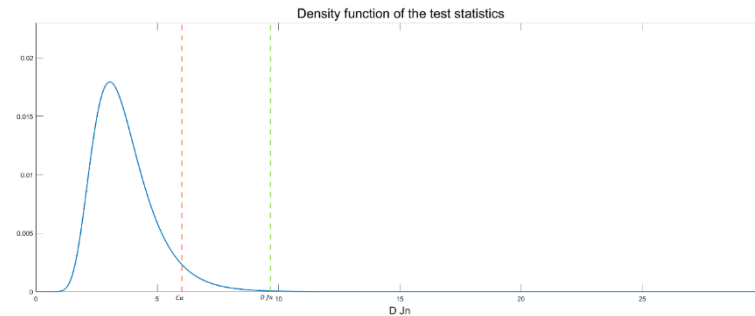
1



2



3



Modelling extreme events

Weighted Maximum Likelihood Estimator

Weighted Maximum Likelihood Estimator (wMLE)

- wMLE is a generalization of MLE concept,
- Let $p = (p_1, \dots, p_n)$, $p_j \geq 0$, $\sum_{i=1}^n p_i = 1$, be a vector of weights assigned to the log-likelihood of each observation, depending on the sample size and possibly on the parameters,
- Weights are assigned through the concept of divergence, which resembles the idea of “distance”, but is not symmetric, $D_{KL}(p) = \sum_{i=1}^n p_i \cdot \log(np_i)$.

Fitting severity distribution by wMLE

- $D_{KL}(p) = \delta$ is a “tilting” parameter calculated based on contamination determined by NPOD,
- Weighted Maximum Likelihood Estimators is the constrained optimization problem

$$\begin{cases} \max_{\theta} \{ \sum_{i=1}^n p_i \cdot \log(f_X(x_i; \theta)) \} \\ D_{KL}(p) = \delta \\ \sum_{i=1}^n p_i = 1 \end{cases}$$

Modelling extreme events

Dependence structure

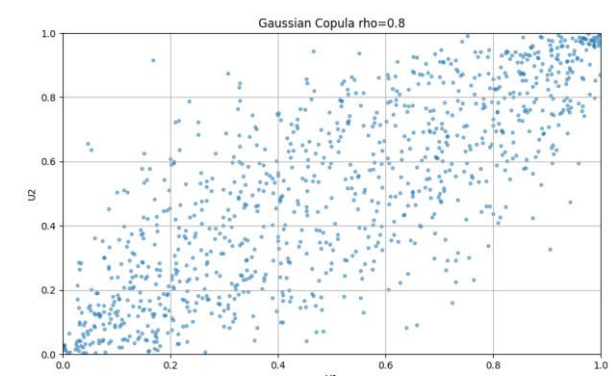
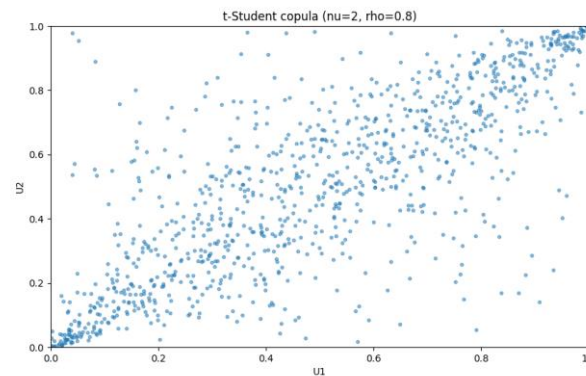
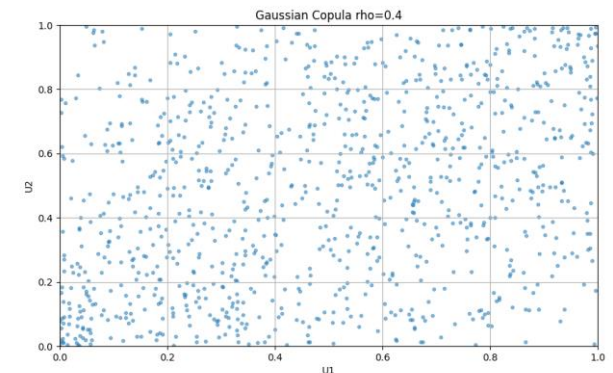
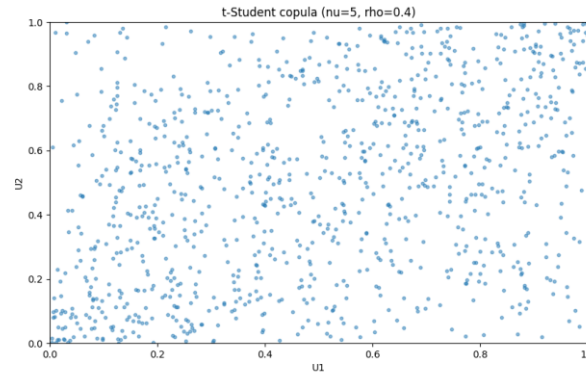
T-Student's copula

$$C: [0,1]^N \rightarrow [0,1]$$

$$C_{\rho, \nu}(u_1, \dots, u_N) = T_{\rho, \nu}(t_{1, \nu}^{-1}(u_1), \dots, t_{N, \nu}^{-1}(u_N))$$

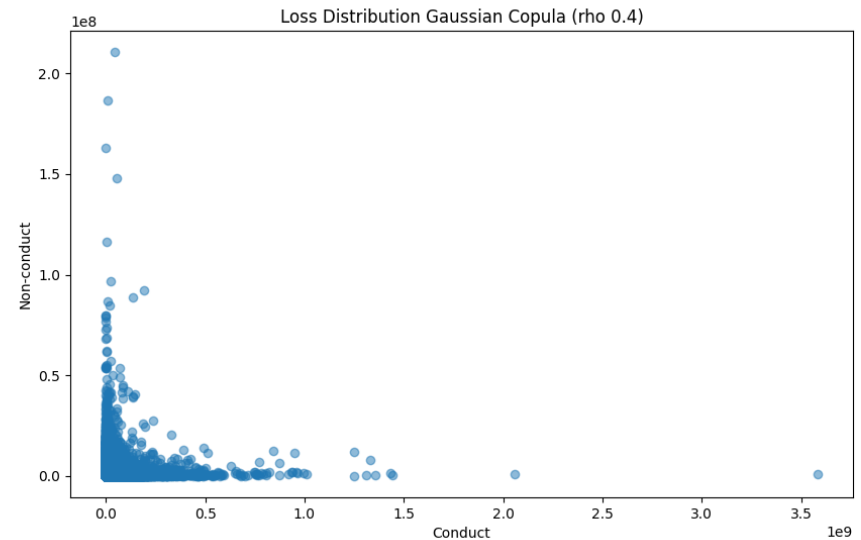
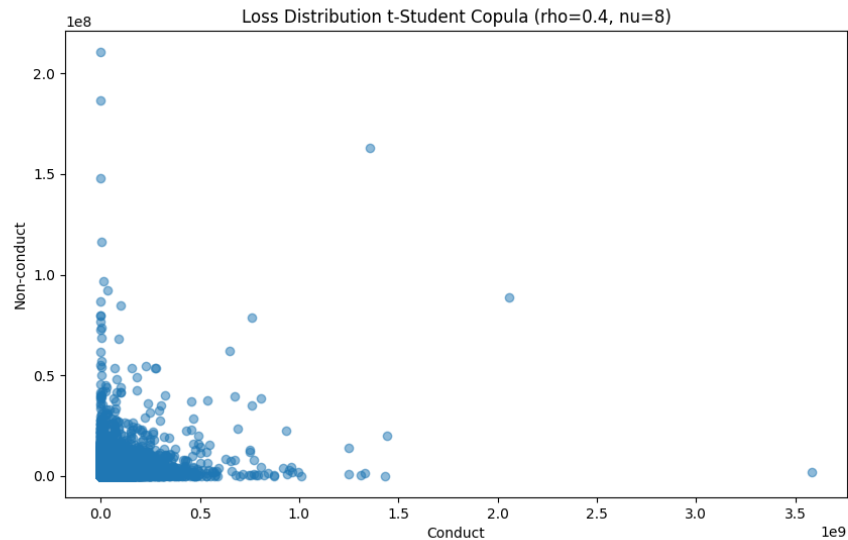
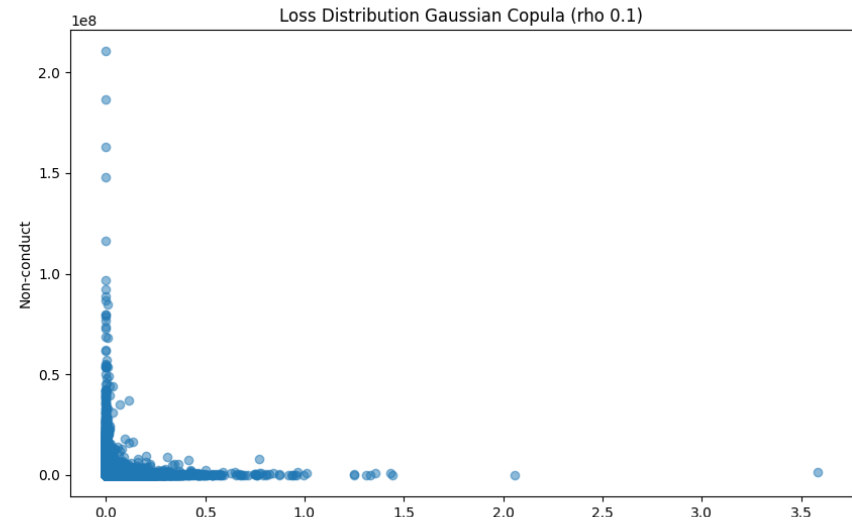
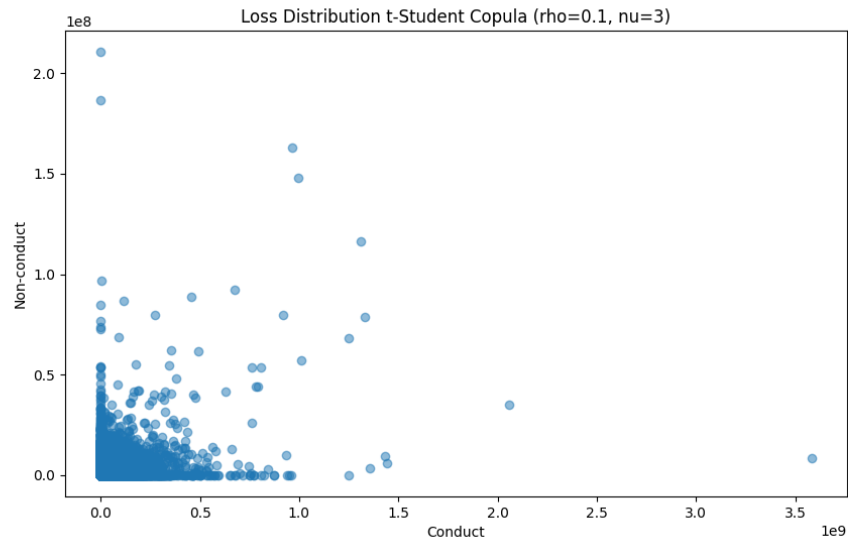
Where $\mathbf{T}_{\rho, \nu}$ is a multivariate Student's t distribution with ρ correlation matrix and ν degrees of freedom and $t_{i, \nu}^{-1}$ is an inverse of standard Student's t cumulative distribution.

- ν controls tail dependency, the lower degrees of freedom parameter the higher dependency between large events.
- ρ is a Pearson correlation, usually calculated in transformation of more robust metrics like Kendall Tau or Spearman's rank correlation.
- T-Student's copula is suitable for modelling dependence structure of extreme events allowing large events to occur simultaneously.



Modelling extreme events

Dependence structure



Modelling Extreme Events

Other challenges in Operational Risk modelling

- ◆ Scarcity of data
- ◆ Model specification
- ◆ Computation time of high frequency simulations
- ◆ High simulation uncertainty
- ◆ Outliers detection
- ◆ Fitting the distribution with outliers using Weighted Maximum Likelihood Estimator (wMLE)
- ◆ Fit assessment including assessment of wMLE fits with weighted goodness of fit tests

References

Non_Parametric Outlier Detector: Bardet J. and Dimby S. (2017), *A new non-parametric detector of univariate outliers for distributions with unbounded support*.

Weighted Maximum Likelihood Estimator: Colombo A., Lazzarini A., Mongelluzzo A., *Weighted Likelihood Estimator for Operational Risk data: Improving the accuracy of capital estimates by robustifying MLE*.

