Beyond Boundaries
astin ▸ afir ▸ iaals
Mexico Colloquia
October 2012

ASTIN Colloquium

1-4 October 2012, Mexico City

Modelling and calibration for non-life underwriting risk: from empirical data to risk capital evaluation

Gian Paolo Clemente
Dept. Mathematics, Finance Mathematics and Econometrics, Catholic University of Milan
gianpaolo.clemente@unicatt.it

Nino Savelli
Dept. Mathematics, Finance Mathematics and Econometrics, Catholic University of Milan
nino.savelli@unicatt.it

Diego Zappa
Dept. Statistical Sciences, Catholic University of Milan
diego.zappa@unicatt.it

THE AIM OF THIS PAPER

By using the well known Danish fire claims data, the target is to explore the performance of the Minimum Distance Approach to fit pure, mixtures and spliced distributions, compared to Maximum Likelihood approach.

This topic is relevant in the actuarial literature in order to analyse the impact of a threshold to separate attritional and large claims in estimating the claim size distribution to be used for risk capital evaluation (premium risk in Solvency II).

Aggregate Claim Amount and claim-size distribution

- Both premium rating and capital requirement for Premium Risk are based on a proper valuation of the aggregate claim amount X for each LoB.

- The aggregate claim amount is well described by a compound process, i.e. the sum of a random number K of random variables Z_j :

$$X = \sum_{j=1}^{K} Z_j$$

- Calibration of the ***claim-size distribution*** (**Z_j**) is a key point in most applications:
  - no standard parametric model seems to emerge as providing an acceptable fit to both small and large claims;
  - the identification of the threshold to separate attritional and large claims is a great task in spliced/mixture distributions;
  - claims are usually posted in the case reserve nearby a “round” number rather than its exact estimation leading to observe probability peaks in the empirical distribution;
  - it could be needed some constraints in parameter estimation (for example, the mean of the fitted distribution must be equal to the mean of the observed data)

Main approaches proposed by actuarial literature

- A two step strategy, based on a ***separate evaluation of attritional and large claims*** is a standard way to describe claim-size distribution:
  - ***Several distributions*** for modelling positive and right-skewed data are proposed in actuarial science (see Klugman et al. (2010))
  - ***Extreme value theory*** and Generalized Pareto distributions are used to describe large claims exceeding a fixed threshold (see McNeil (1996), Embrechts, Kluppelberg, Mikosch (1997))

- Other approaches are based on mixtures and combined distributions:
  - **Frigessi et al** (2002) propose a weighted mixture model based on a GPD and on a light-tailed distribution
  - **Cooray & Ananda** (2005) combine LogNormal and Pareto distributions by fixing the proportion of large claims
  - **Teodorescu, Vernic** (2007), **Teodorescu, Panaitescu** (2007), **Vernic et al.** (2009) provide different mixtures based on Exponential-Pareto, Weibull-Pareto and LogNormal-LogNormal.
  - **Scollnik** (2007) expands Cooray & Ananda paper by estimating the threshold directly by data
  - **Pigeon and Denu**it (2011) extend the LogNormal-Pareto model by assuming a random threshold (Gamma or LogNormal distributed)
  - **Nadarajah and Bakar** (2012) try to improve fitting to Danish Data by using a composite distribution based on a LogNormal and various distributions for large claims. They assume the LogNormal-Burr as the best one for Danish Data.

⇨ All papers use ***maximum log-likelihood (ML) approach*** to estimate parameters.

⇨ Most of them use the public Fire Danish losses database to test the performance of their own method

Minimum Distance Estimation Approach (MDA)

- The *MDA* method (***Parr*** (1985), ***Basu et al.*** (2011)) consists in solving the general unconstrained problem

$$\boxed{\min_{\theta} \quad d(\hat{F}_n(y), F_Y(y;\theta)) \quad Y \in \Re_Y}$$

where

$\{y_1, y_2, \ldots, y_n\} \in Y_n$ Random Sample of Y~W(y; θ)

$\hat{F}_n(y) = \frac{\sum_{j=1}^{n} I_{\{y_j \leq y\}}}{n}$ Empirical Distribution of Y_n

$F_Y(y;\theta) = P(Y \leq y)$ Cdf of Y_n

$d(\cdot)$ Appropriate Loss Function

Examples of d

$$d(A(x), B_\theta(x)) := \begin{cases} KS: \ \sup|A(y) - B_\theta(y)| \\ MH: \ E[|A(y) - B_\theta(y)|] \\ AD: E[(A(y) - B_\theta(y))^2 \cdot w_y] \end{cases}$$

Kolmogorov-Smirnov
Manhattan
Anderson-Darling

If w_y=1, AD yields the CvM distance

- If $F_Y(y;\theta)$ is the *"true" distribution* the MDA unconstrained estimator **has good statistical properties:**
  - it is strongly consistent, i.e. $P\left\{\lim_{n\to\infty}\hat{\theta}=\theta\right\}=1$
  - depending on which d(·) is used (i.e. for KS or MH), it is robust to influence values
  - if **CvM** loss function is used, the estimators are asymptotically normal

- Furthermore, the algorithm can be simply redesigned conditional to $Y \in \Re_c \subseteq \Re_Y$

  The general constrained mixed problem can be stated as follows:

$$\min_{\theta|Y\in\Re_{c\,1}} d_1(\hat{F}_n(y), F_Y(y;\theta))$$

$$s.t. \quad \min_{\theta|Y\in\Re_{c\,2}} d_2(\hat{F}_n(y), F_Y(y;\theta)) < \delta$$

where δ is the maximum error, d_1,d_2 and $\Re_{c1}, \Re_{c2}$ are respectively arbitrary discrepancy function and constrained regions.

- Having chosen a distribution:
  - ML: aims at estimating parameters such that they include the maximum likelihood coming from the sample. The estimates drive the shape of the theoretical distribution.
  - MDA: aims at finding parameters such that is max the fitting to the empirical Cdf. The estimates are driven by the shape of the empirical distribution.

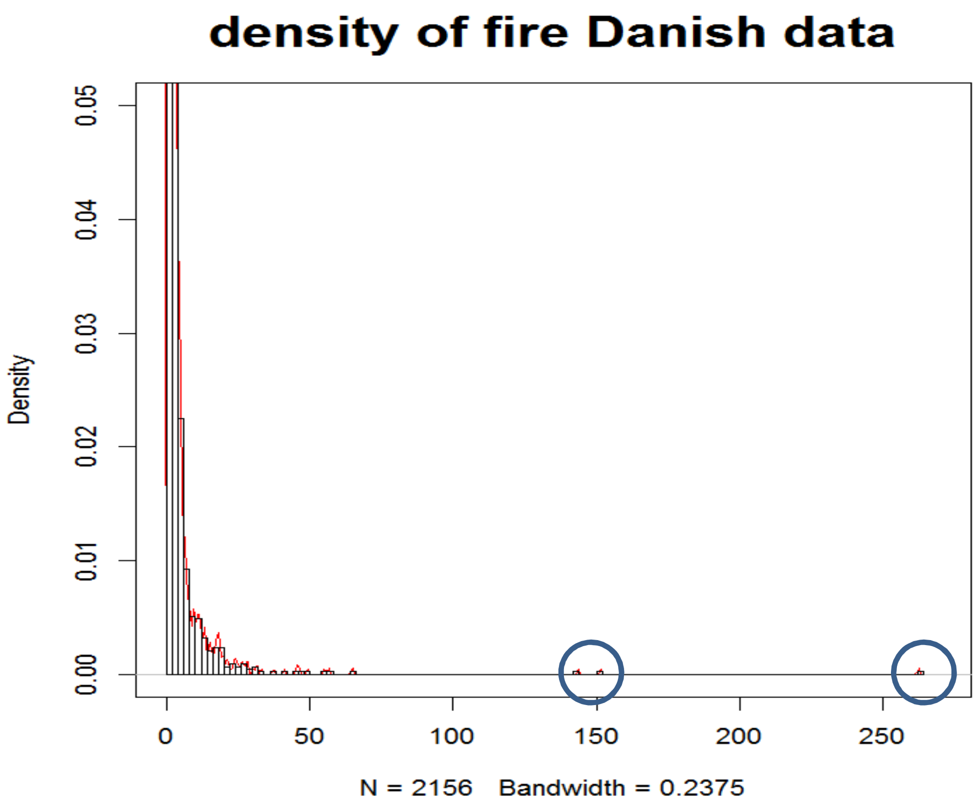
- **MDA Advantages:** being data driven it tries to adapt estimates to the natural shape of the empirical distribution and moreover it is easy to implement even in a spreadsheet.
  It overcomes the problem of not convergence of ML optimization algorithm due to the existence of local maximum in the likelihood function (see e.g. estimation of parameters of mixtures).
  Distance measures may be easily modified to include weights.

- **MDA Drawbacks:** asymptotic distributions of estimators are not easy to be derived in closed form. *Bootstrap procedures are often needed*. If no weights are used and dataset is large, information from relevant but scarce data (e.g. extremes) are lost by the procedure.

Case study: the Fire Danish Data

- Consider the well-known **fire Danish data-set (1980-1990)** (see McNeil (1997)).
- Some papers report results based on ML approach considering all the data. As originally proposed by McNeil we have fitted distributions only on losses <u>over 1 million of DKK.</u>

Main Characteristics	
N. Obs	2156
Mean	3.397
St. Dev	8.53
CV	2.51
Skewness	18.75
Kurtosis	483.46
Min	1.003
1^{st} Quartile	1.331
3^{rd} Quartile	2.973
Max	263.30



Comparison of MDA vs ML
(1) *Pure distributions*

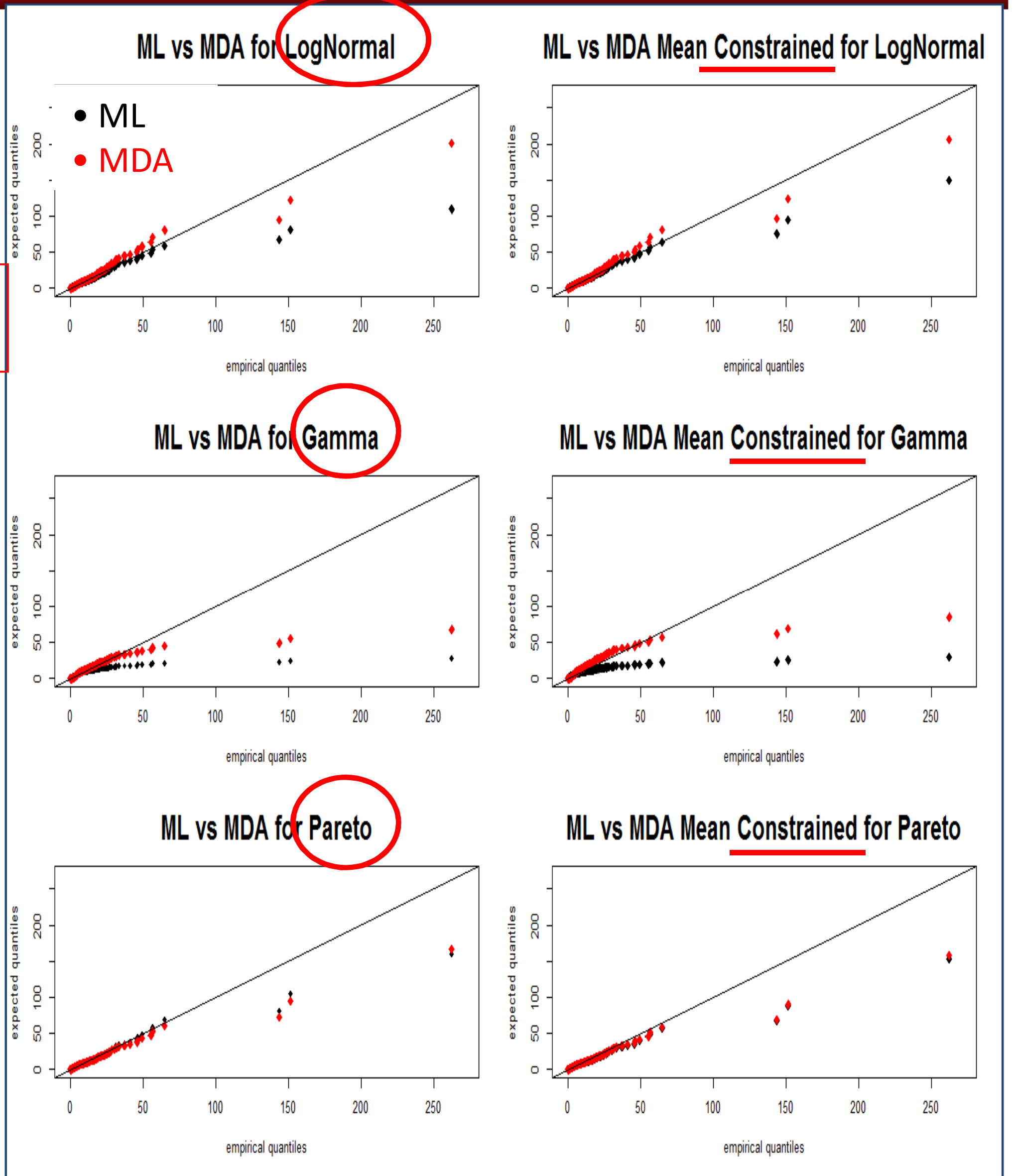
- **Modified AD Loss function** has been considered:

$$d\left(\hat{F}_n(y), F_Y(y;\theta)\right) := \sum\left(\hat{F}_n(y) - F_Y(y;\theta)\right)^2 \cdot w_y$$

- MDA has been calibrated in order to improve fitting on the tails **by introducing increasing weights (w_y=y).**

- The goodness of fit has been tested also by introducing the constraint on observed mean:

$$E(Y) = \frac{1}{n}\sum y_i$$



Comparison of MDA vs ML

(2) *Mixtures (LogNormal-Pareto and Pareto-Pareto)*

- Analogously to results available in the literature, we have fitted data by a **mixture of distributions:**

$$F_Z(z) = \pi F_{Z_1}(z) + (1-\pi) F_{Z_2}(z) \quad \text{with } 0 < \pi < 1$$

by assuming

$$Z_1 \sim LogN(\mu, \sigma^2), \;\; Z_2 \sim Pa(\alpha, \theta)$$

- **LogNormal-Pareto mixture**

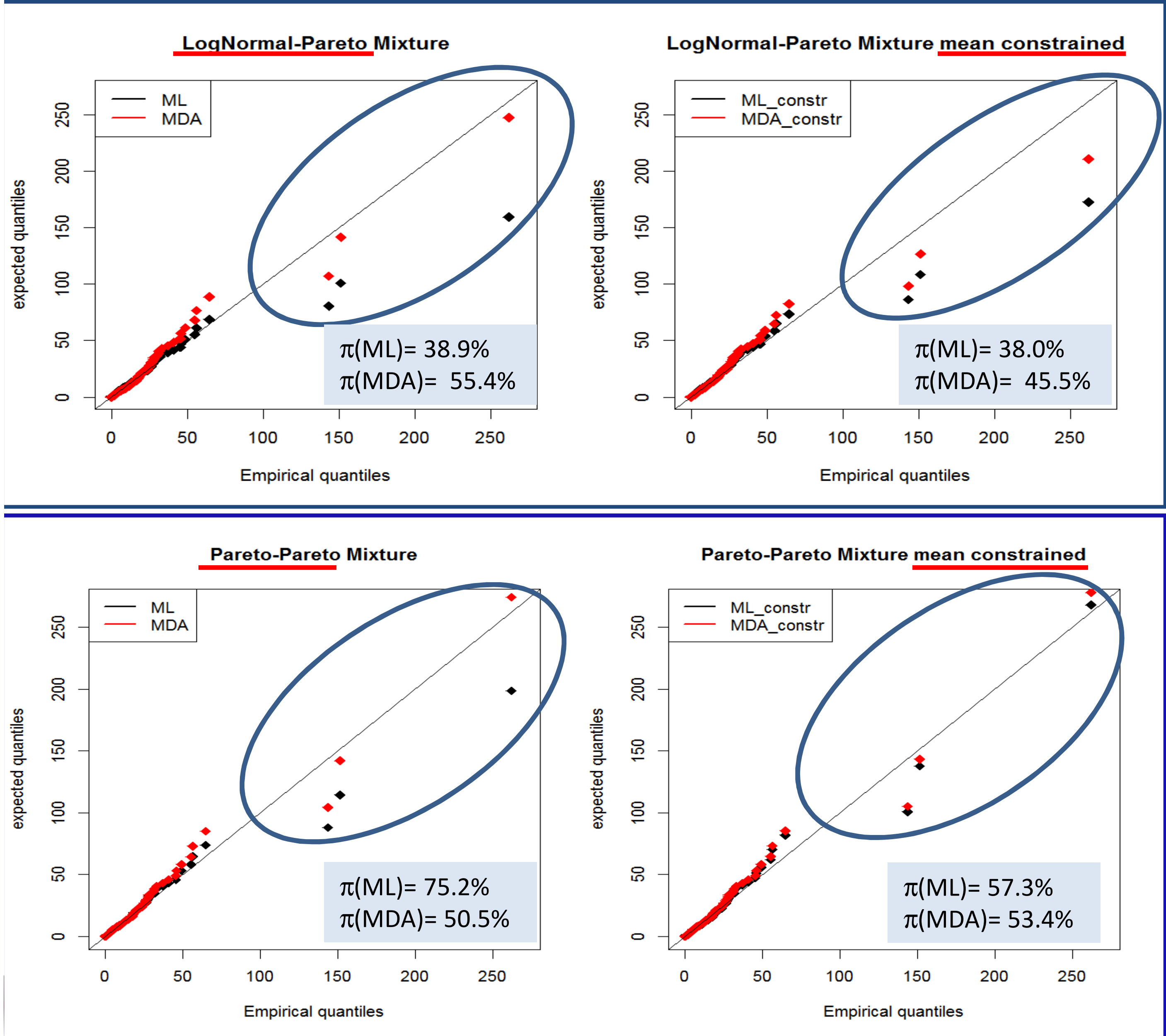
or

$$Z_1 \sim Pa(\alpha_1, \theta_1), \;\; Z_2 \sim Pa(\alpha_2, \theta_2)$$

- **Pareto-Pareto mixture**

- Both mixtures have been calibrated with and without the constraint:

$$\boxed{E(Z) = \pi E(Z_1) + (1-\pi) E(Z_2) = \frac{1}{n} \sum y_i}$$



In both cases, even without the constraint on the mean, the MDA approach seems to perform better than ML in fitting extremes values. Moderate (positive) bias exists in fitting attritional losses.

ML fit of the tail improves when the constraint on the mean is imposed.

Spliced distributions (MDA)

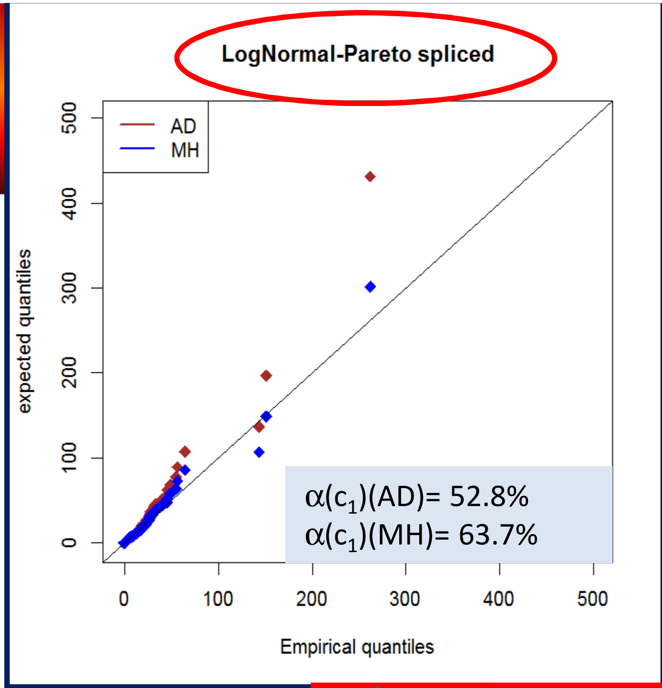
- In the literature a modern approach is based on the estimation of spliced distributions. The corresponding probability density function for a random variable Z with domain (c_0,c_2) is defined as:

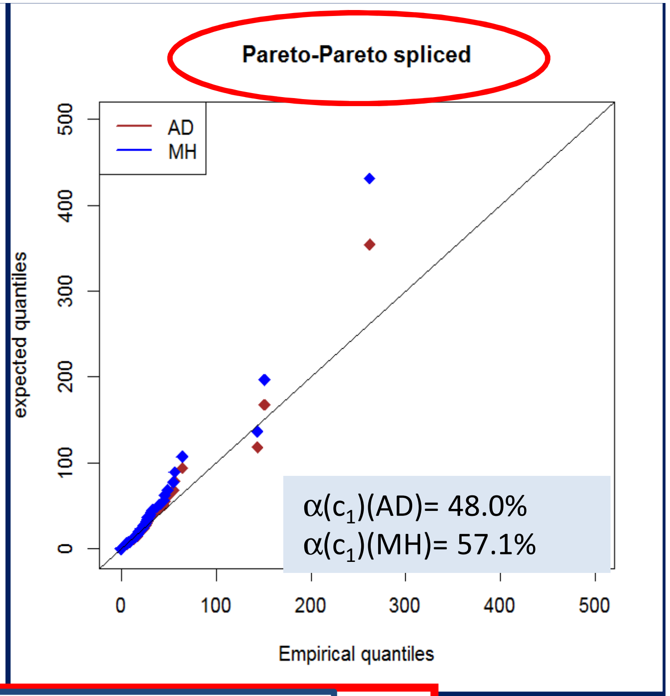
$$\boxed{f_Z(z)=\begin{cases}\alpha(c_1)\cdot f^*_{Z_1}(z) & c_0<z<c_1\\ (1-\alpha(c_1))\cdot f^*_{Z_2}(z) & c_1\le z<c_2\end{cases}}$$

  - $\alpha(c_1)$ is the percentile corresponding to the quantile c_1
  - c_i is the limit of the domains
  - f* is a truncated probability density function

$$f^*_{Z_i}(z)=\frac{f_{Z_i}(z)}{\int_{c_{i-1}}^{c_i} f_{Z_i}(z)dz}\quad \text{for}\quad i=1,2$$

Further conditions must be imposed if continuity and differentiability at the knots are needed

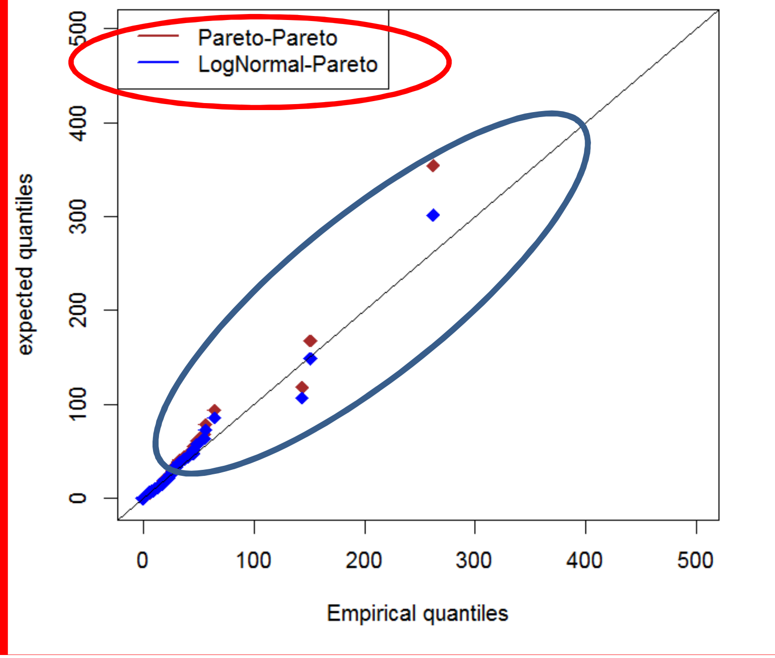




Figures show the *LogNormal-Pareto spliced (L-P)* and the *Pareto-Pareto spliced (P-P)* fitted by MDA.

MDA provides good results by using MH for *L-P* and AD for *P-P*

L-P(MH) vs P-P (AD)



Conclusions

- MDA may assure in general a better fit than ML due to its property to adapt the distribution to the data. It is not suitable if inference on the parameters is needed.

- A good fit of extreme values may be assured when weights are used. The *choice of both optimal weights and appropriate loss function* at the moment represents an element of subjectivity.

- Moreover, it could be emphasized that the good fit on the tails, derived by MDA, could assume a great importance in risk capital evaluation where extreme percentiles of aggregate claim amount are considered (as $VaR_{99.5\%}$ or $TVaR_{99\%}$ in Solvency II and SST).

- Furthermore, we have applied MDA also to real MTPL datasets (see Appendix at this regard).

- The good performance of MDA have been confirmed as well as the evidence that in general severity distributions are often not well described by “pure” distributions.

- Finally, when the entire shape of empirical distribution is considered, the presence of peaks, that are usually given by case reserve evaluation, can produce a worse fitting. A Normal-Normal spliced distribution on Log(data), calibrated using a MDA approach, provides the best fitting on the empirical MTPL data-set.

- Further researches: an algorithm for the choice of optimal weights in MDA, the use of mixed distribution, bootstrap procedures in order to obtain estimators distributions, sensitivity on risk capital of alternative approaches.

Main references

- **Basu A., Shioya H., Park C**. (2011), *Statistical inference: the minimum distance approach*, CRC Press, 2011;
- **Cooray K., Ananda M M A** (2005), *Modelling actuarial data with a composite LogNormal-Pareto*, Scandinavian Actuarial Journal 5, 321-334;
- **Denuit M., Pigeon M.** (2011), *Composite LogNormal-Pareto model with random threshold*, Scandinavian Actuarial Journal 3, 177-192;
- **Frigessi A., Haug O, Rue H** (2002), *A dynamic mixture model for unsupervised tail estimation without threshold selection*, Departement of Mathematical Science, Norway;
- **Klugman S., Panjer H.H., Wilmot G.E.** (2010), *Loss Models: From Data to Decisions*, John Wiley & Sons;
- **McNeil A** (1997), *Estimating the Tails of Loss Severity Distributions using Extreme value Theory*, ASTIN Bullettin 27, 177-137;
- **Nadarajah S, Bakar A A S** (2012), *New composite models for the Danish-fire data*, Accepted and to appear on Scandinavian Actuarial Journal;
- **Parr W.C., Shucany W. R.** (1980), *Minimum Distance and Robust Estimation*, Journal of the American Statistical Association, Vol 75, n. 371, 616-624;
- **Scollnick D P M** (2007): *On a composite LogNormal-Pareto model*, Scandinavian Actuarial Journal 1, 20-33;

APPENDIX

An analysis of a MTPL dataset

An empirical distribution for MTPL

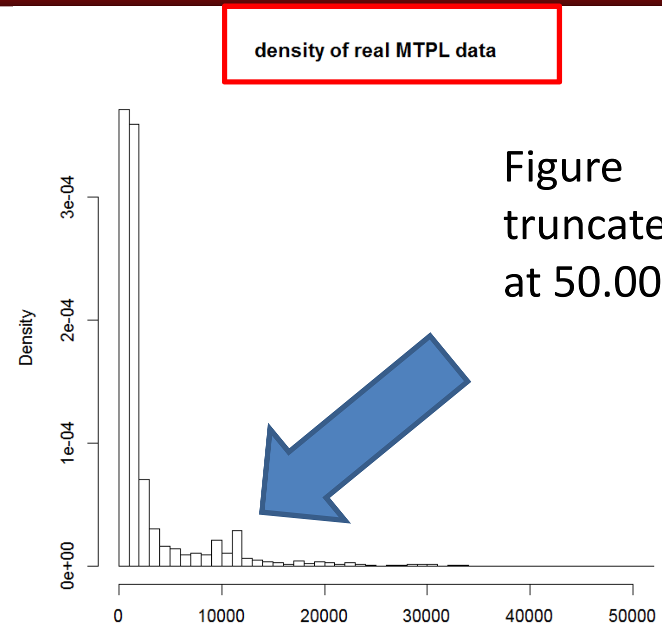
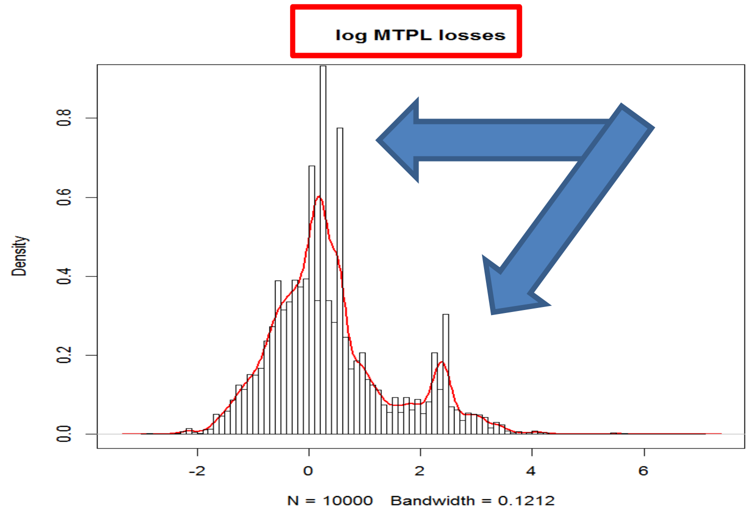


Figure truncated at 50.000



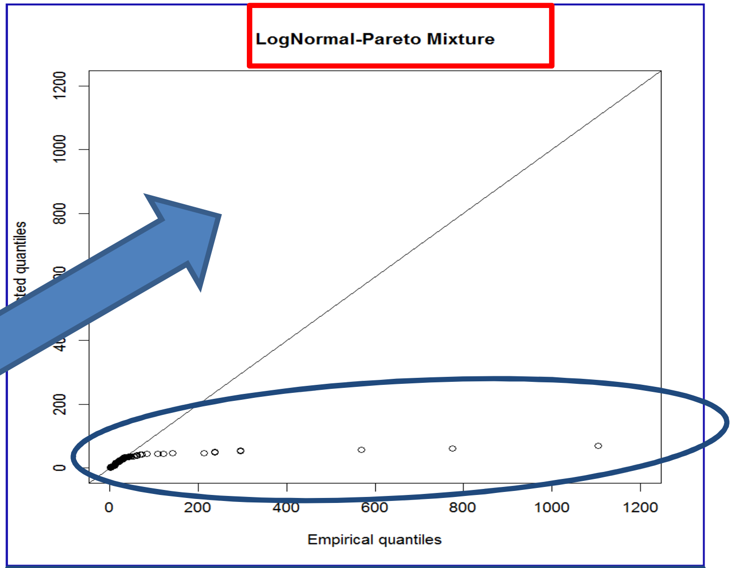
Main Characteristics	
N. Obs	11568
Mean	3487.75
St. Dev	16875.04
CV	4.84
Skewness	43.44
Kurtosis	2410.09
Min	50.00
1st Quartile	544.80
3rd Quartile	2278.70
Max	1.105 mln

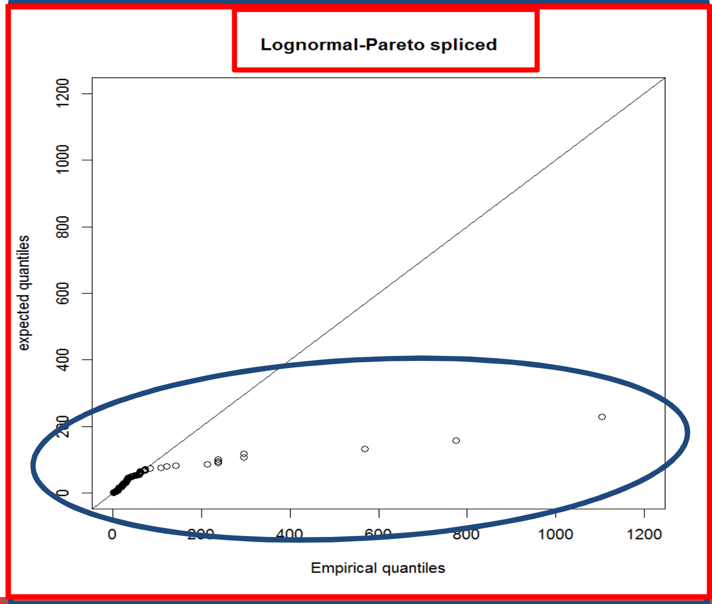
Peaks around 1.000, 10.000, 12.500 € can be observed

MTPL data

(1) *Pure, Mixture and Spliced distributions*

- Considering now MTPL data previously described, we obtain a poor fitting by using (with a maximization based on either ML or MDA):
  - Pure distributions (LogNormal, Pareto, Gamma, etc have been tested)
  - Mixture distributions (see LogNormal-Pareto in the upper figure. Gamma-Pareto and LogNormal-Gamma, here not reported, have been applied too)
  - Spliced Distribution (see Logormal-Pareto)

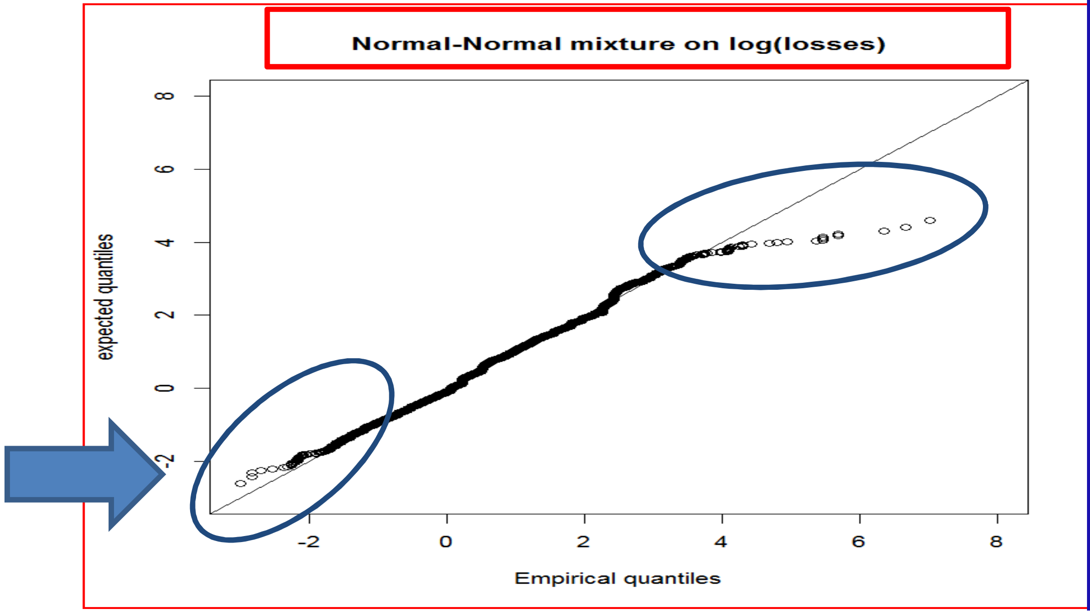




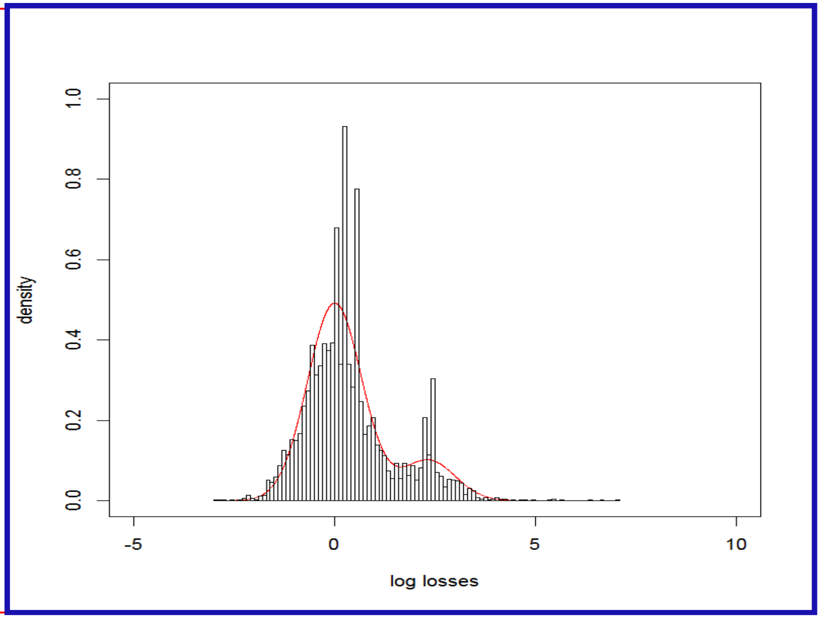
MTPL Log(data)
(2) *Mixtures*

- Pure, Mixtures and Spliced distributions provide a poor fit of MTPL data.
- Considering now the empirical distribution of the log of losses, a Normal-Normal Mixture have been applied.
- Only MDA results are here reported

q-qplot

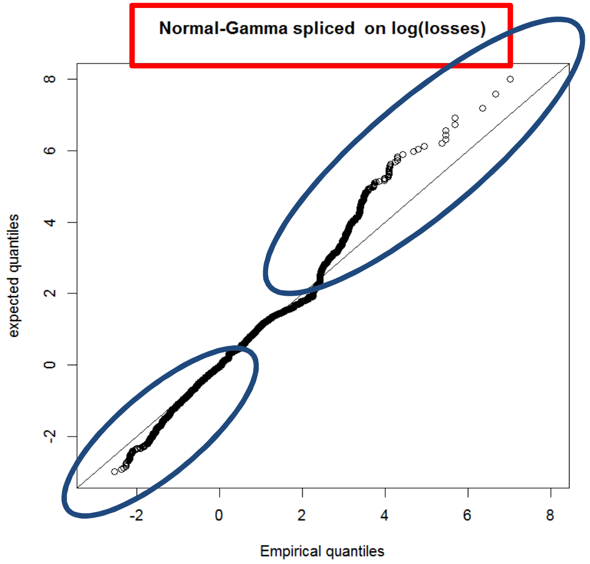


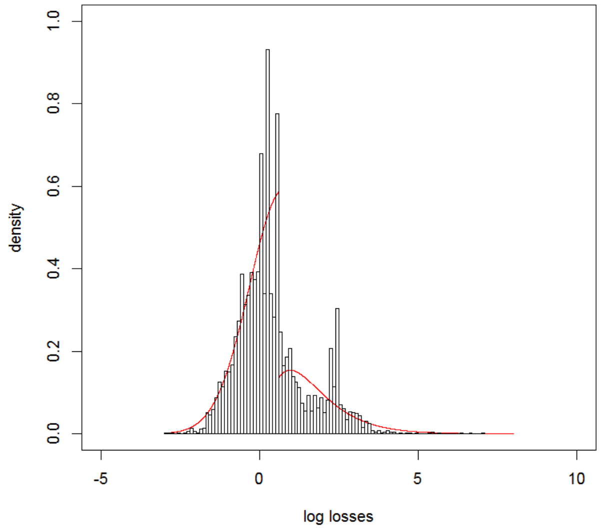
p.d.f.



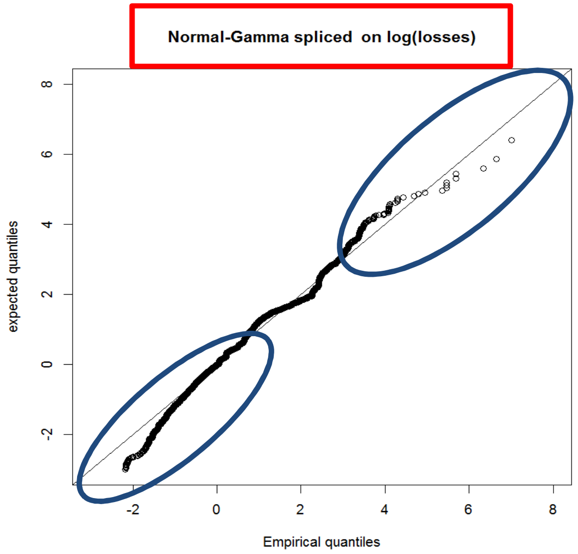
MTPL

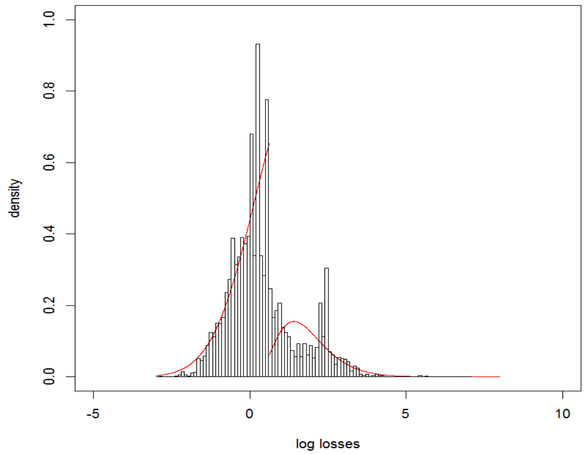
(3) *Normal-Gamma Spliced on Log(Losses)*





Loss Function: AD





Loss Function: KS

MTPL

(4) *MD Approach: Normal-Normal Spliced on Log(Losses)*