Modelling Premium Risk for Solvency II: from empirical data to risk capital evaluation

A Revisited Minimum Distance Approach for modelling severity claims

(this draft, 08/06/2013)

Clemente Gian Paolo [*,1] , Savelli Nino [*] , Zappa Diego [**]

[*] *Department of Mathematics, Finance and Econometrics,* [**] *Department of Statistical Sciences*

Università Cattolica del Sacro Cuore - Milan, Italy

Abstract

Solvency II will introduce economic risk-based solvency requirements for insurance companies across all European Member States for the first time. The new solvency requirements will be more risk-sensitive than in the past, allowing a better coverage of the real risks run by any particular insurer. Focusing only on technical risk, that has usually the greatest impact on the capital requirement for Non-Life insurers, particular attention should be paid to identify accurately the distribution of aggregate claim amount. The reason is that collective risk models are usually applied with a separate evaluation of frequency and severity distribution. For the latter component most of the papers use Maximum Likelihood methods to estimate parameters. However, in the practice it is common to observe unacceptable fit to both small and large claims of severity distribution, because of either the chosen model or the non-robustness of the estimation procedure methodology. To overcome this issue we propose to estimate parameters via Minimum Distance Approach (MDA). Focusing on a common case study (the Danish fire claims provided by McNeil) extensively studied in actuarial literature, we show that MDA has superior capability to fit insurance data.

Keywords Claim-size distribution, Mixture and spliced, Minimum Distance Approach, Capital requirement.

[1] Corresponding Author: e-mail gianpaolo.clemente@unicatt.it

Introduction

Insurers need to know the expected level of claims they will face within future years and the percentiles of the aggregate claim amount distribution so that they can estimate the level of pricing and of the required capital. The usual way of doing that is by an aggregate claim model. It aims to fit both a probability distribution to the number of claims and a distribution to the size of the individual claims received. Combining them into a composite Poisson process the distribution of the aggregate amount of claims is obtained often through simulation porcedures. The resulting aggregate distribution is then used to estimate percentiles and hence the level of solvency capital required in future years.

Focusing only on the severity distributions, it is a common assumption to consider costs originating from a mix of small and large claims. At this regard many works show that standard parametric model usually does not provide an acceptable fit to both small and large claims.

Univariate distributions, despite proving a good overall fit, might indeed fail in fitting tails, so that a two step strategy, based on a separate evaluation of attritional and large claims is the standard alternative to describe claim-size distribution (at this regard see also the Non-Life capital requirement evaluation defined by Swiss Solvency Test). Several distributions for modelling positive and right-skewed data have been proposed in actuarial science (see Klugman et al. [7]) as well as extreme value theory is often used to describe large claims exceeding a fixed threshold (see Embrecht [5]).

To test the performances of the many alternatives sketched above, McNeil [10] applied different distributions in order to assure a good fitting for the Danish fire Insurance data that is a well known database in actuarial literature and often used to test the efficacy of new proposals. Both classical distributions (such as LogNormal and Pareto) and GPD have been applied. The main drawback is the difficulty to estimate the threshold in order to separate large losses. 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 as defined in premium risk by Solvency II.

Viable approaches are based on the use of mixture and composite models. The former was applied by Frigessi et al. [6] with a weighted mixture model based on a GPD and a light-tailed distributions in order to avoid threshold selection.

Recent findings by Cooray and Ananda [2] show instead that composite models can give better fits. Their application to Danish data is based on a composite Lognormal-Pareto model derived by assuming a scaled LogNormal distribution up to an unknown threshold value (equal to the parameter θ of the Pareto distribution) and a scaled two parameters Pareto distribution for large values. The resulting distribution is similar to a LogNormal density with a thicker upper tail.

Scollnik [16] improved the composite LogNormal-Pareto model by using mixing weights as additional parameters for each piecewise function, replacing the constant weights applied by Cooray and Ananda, and by using also a GPD distribution for large claims.

Pigeon and Denuit [14] assume that a unique threshold value applied to all the claims may appear quite unrealistic, and they improve the model by assuming a random threshold based on a Gamma or a LogNormal distribution. Finally Nadarajah and Bakar [11] tries to improve the fitting to Danish data by comparing new composite models based on a LogNormal and various distributions for large claims. They find a good fitting with the composite LogNormal-Burr.

The estimation procedure of the parameters of all the models reported above is based on a Maximum Likelihood (ML) methodology.

Aim of the paper is to investigate the efficacy of the minimum distance approach (MDA) for parameters estimation. The choice of this fitting method grounds on the fact that in actuarial problems the issue is primary to obtain a good fitting (see e.g. the many qq or pp-plots used by several authors to show the adequacy of their fitting) and secondarily to check the quality of the fitting by classic inferential procedures. MDA, being a data driven methodology, may obtain, in some cases, a better fitting than Maximum Likelihood.

Using Danish fire claims data, the target is to explore the performance of Minimum Distance Approach (MDA) to fit pure, mixtures and spliced distribution compared to Maximum Likelihood Approach. In particular we extend the classical MDA approach by introducing alternative loss functions and weights on the empirical data.

Mixtures, Composite, Spliced distributions

In this Section we give a brief review of the most interesting model used in actuarial literature that will be compared later to test the efficacy of MDA in fitting different type of distributions. Standard univariate distributions are well known in the literature and will not be reported in detail.

A first way where long-tailed and/or skewed distribution arise naturally is through a mixture of different distributions.

Consider $X_1, \dots, X_n$ random variables with probability density function (p.d.f.) $f_1(x), f_2(x), \dots, f_n(x)$ and mixing weights $\pi_1, \pi_2, \dots \pi_n$ with $\sum_{i=1}^{n} \pi_i = 1$. The p.d.f. of the mixture of $X_1, \dots, X_n$ is:

$$f(x) = \sum_{i=1}^{n} \pi_i \cdot f_i(x)$$

Several applications of a mixture can be found in actuarial literature (see [8] for a first example of loggamma and gamma mixture as a viable model for claim distribution). Focusing on Danish data, Frigessi et al. [6] applied a mixture of a Weibull and a GPD distribution.

An alternative to a mixture is through a Combined distribution. It is defined as:

$$f(x) = \begin{cases} c \cdot f_1(x) & 0 < x \leq \theta_1 \\ c \cdot f_2(x) & \theta_1 < x \leq \theta_2 \\ & \dots \\ c \cdot f_n(x) & \theta_{n-1} < x \leq \infty \end{cases}$$

where θ_i for i=1,…, n-1 are thresholds and c a normalizing constant. In general composite models are used for only a couple of r.v.. Cooray and Ananda [2] use a Combined LogNormal- Pareto distribution with $f_1(x)$ is a LogNormal(μ,σ) density defined for x greater than zero and $f_2(x)$ is a Pareto density with only one parameter and a lower truncation point at θ. In this case the normalizing constant c necessary to assure that *f(x)* is a proper density function may be found as follows:

$$\int_0^{\infty} f(x)dx = \int_0^{\theta} c \cdot f_1(x)dx + \int_{\theta}^{\infty} c \cdot f_2(x)\, dx = 1$$

Since Pareto is defined over the range $[\theta, \infty)$, c can be derived by the equation:

$$\int_0^{\theta} c \cdot f_1(x)dx + c = 1$$

When $f_1(x)$ is a LogNormal

$$c = \frac{1}{1 + \int_0^{\theta} f_1(x)dx} = \frac{1}{1 + \Phi\left(\frac{\ln(\theta) - \mu}{\sigma}\right)}$$

The authors impose then continuity and differentiability conditions:

$$f_1(\theta) = f_2(\theta)\,, f_1'(\theta) = f_2'(\theta)$$

which imply the constraints: $\ln(\theta - \mu) = \alpha\sigma^2$ and $\exp(-\alpha\sigma^2) = 2\pi\alpha^2\sigma^2$. The aim is to obtain a smooth probability density function, but this choice limit the flexibility of the overall distribution since the parameters reduce from four to two (as observed by Scollnick in [16]).

So the approach has been extended by Scollnick introducing a Composite Lognormal-Pareto defined as:

$$f(x) = \begin{cases} r \cdot \dfrac{1}{\Phi\left(\dfrac{\ln(\theta)-\mu}{\sigma}\right)} f_1(x) & 0 < x \leq \theta \\ (1-r) \cdot f_2(x) & \theta \leq x \leq \infty \end{cases}$$

where r is a mixing weight and $1/\Phi(\cdot)$ assures that $\int_0^\theta \dfrac{1}{\Phi\left(\dfrac{\ln(\theta)-\mu}{\sigma}\right)} f_1(x)dx = 1$

The mixing weight r is again obtained by imposing continuity and differentiability conditions, reducing to three unknown parameters (θ,σ,α). An alternative extension has been obtained in the same framework by assuming a Composite LogNormal-GPD. Notwithstanding these models benefit of the property of having fat tails they provide a poor fitting for upper quantiles of Danish data. It has been noticed an overestimation of empirical quantiles with the composite LogNormal-Pareto from 90th percentile and with the LogNormal-GPD from 99th.

Further extension has been provided by Pidgeon and Denuit by assuming the threshold value as a realization of a positive random variable. They derive closed formulae by assuming the threshold being random and Gamma or LogNormal distributed. Results show a reduced overestimate of extreme quantiles. Finally, Nadarajah and Bakar show, without providing numerical results (only graphic plots are reported in [11]), that a Composite LogNormal-Burr can assure a better fitting, despite central quantiles appears to be overestimated. To estimate parameters all the mentioned approaches made use of the Maximum Likelihood approach.

A third family of distribution that we will explore in the next session is the spliced distribution. A n component spliced distribution has a density function that can be expressed as follows:

$$f(x) = \begin{cases} \pi_1 \cdot f_1^*(x) & c_0 < x \leq c_1 \\ \pi_2 \cdot f_2^*(x) & c_1 < x \leq c_2 \\ \quad\quad\quad\quad \dots & \\ \pi_n \cdot f_n^*(x) & c_{n-1} < x \leq c_n \end{cases}$$

where $\sum_{j=1}^{n} \pi_j = 1$, and $f_j^*(x)$ are truncated distribution over the domain $(c_{j-1}, c_j]$ for j=1,…,n.
That implies $\int_{c_{j-1}}^{c_j} f_j^*(x)dx = 1$.

Previous composite functions could be reinterpreted as a two-component spliced model. This approach (as the composite models) allows a direct estimation of the threshold by defining attritional and large claims as produced by different random variables. Furthermore avoiding continuity and differentiability conditions, the number of unknown parameters are not restricted.
This choice can improve the fitting with greater degrees of freedom but, at the same time, the optimization algorithm could fall in some convergence problems.
Finally, further hints come from studying a smoothed distribution that could be obtained without imposing previous conditions but through some smoothing methodology (see Charpentier and Oulidi [2] for a beta kernel approach).

Minimum Distance Approach

As already said above to estimate parameters of distributions, we will use a minimum distance approach (MDA) (see Basu [1]). This approach has not yet extensively explored in actuarial literature. The estimation by an appropriate minimum distance method is a natural idea in statistics. In general the aim is to estimate the parameters of a parametric cumulative density function (cdf) by minimizing the distance between the empirical cdf and a model.
We have chosen to apply this method, opposite to the most common Maximum Likelihood method, because ML aims at finding the parameters of a distribution such that is maximum the representativeness (in the case of discrete r.v. it is equal to the probability) of a sample through the estimated model with respect to the unknown population. By contrast when the aim is to find the distribution that best suits the data that almost sure represent the population (i.e. the sample is so large that the sample itself is somehow the population) and less importance is posed to inferential issues, data driven approaches, as MDA, may have superior performance.
In our context, we assume to be more interested to an approach that adapt estimates to the natural shape of the empirical distribution rather than to inferential issues as in ML. In general MDA consists in solving a general unconstrained problem

$$\min_\theta d\,(F_n(\boldsymbol{y}), F_Y(\boldsymbol{y}, \theta)) \qquad (1)$$

where $d(\cdot)$ is an appropriate loss function, $F_n(\boldsymbol{y}) = \frac{\sum_{j=1}^{n} I_{\{y_j \le y\}}}{n}$ the empirical cdf and $F_Y(\boldsymbol{y}, \theta)$ the cdf of the theoretical distribution, Y, with parameters θ.
If it exists a solution $\hat{\theta} \in \Theta$ to (1) then $\hat{\theta}$ is called the minimum distance estimate of θ.
Several loss functions can be chosen in (1): let $A(x)$ and $B_\theta(x)$ be continuous functions, examples of $d(\cdot)$ are

$$d\big(A(x), B_\theta(x)\big) := \begin{cases} KS: \quad sup\,|A(y) - B_\theta(y)| \\ MH: \; E[|A(y) - B_\theta(y)|] \\ MQ: \; E\left[\big(A(y) - B_\theta(y)\big)^2\right] \\ AD: \; \sum \big(A(y) - B_\theta(y)\big)^2 w_y \end{cases}$$

i.e. the Kolgomorov, Manhattan, Euclidean (Cramer – von Mises), Anderson Darling distances, respectively. It is noteworthy that depending on which distance is used, we could have asymptotical property of the estimators (see Parr [13] for details).

A generalization of the approach is by using a weighted distance of the form:

$$d(F_n(x), F_\theta(x)) = \sum_{i=1}^{n} \big(F_n(y_i) - F_Y(y_i; \theta)\big)^q y_i^p \tag{2}$$

where $q > 0$ and $p \geq 0$. (2) extends classical distances by using parameter q as the power of the distance. Since in our context we have to deal with positive skewed distribution and large losses have significantly greater importance than small claims, furthermore, we introduce weights, represented by the empirical values y_i, allowing to give a different importance to the fitting of attritional and large claims. The larger is p, the larger will be the weight given to large values. Note that this choice could be appropriately modified if we consider symmetric distribution.
It means we can solve previous relation by searching also parameters (p, q). The parameters p and q can represent two additional elements priorly chosen or estimated by an iterative algorithm in order to improve the fitting of the parametric distribution to the empirical one.
For particular choices, we come back to classical loss functions. For example, for $q = 2$ and $p = 0$, the approach leads to Cramer Von Mises loss distance, while with $q = 2$ and $p = 1$ we come back to Anderson Darling distance.

To keep distances within known family we will fix q=1, 2 by letting the power p to vary. Since for different q, p distances are not fully comparable, to compare the fitting for different combination of q, p we have used the distance

$$D = \left(\sum_{i=1}^{n} \left(y_i - F_Y^{-1}(F_n(y_i); \hat{\theta}_{p,q})\right)^2\right)^{1/2} \tag{3}$$

i.e. the Eucledian distance between the data and the quantiles obtained from of F_Y with $\hat{\theta}_{p,q}$ the corresponding parameter estimates. Note that a straightforward alternative to (2) (not explored in this paper) might be

$$d\left(y_i, F_\theta^{-1}(F_n(y_i))\right) = min_\theta \sum_{i=1}^{n} \left(y_i - F_Y^{-1}(F_n(y_i); \hat{\theta}_{p,q})\right)^q y_i^p \tag{4}$$

The solution to (2) and (4) are not necessarily the same. We will focus on (2) for a matter of correspondence with the literature.

Empirical analyses

In this Section, we report the main results of an application of relations (1) and (2) to the Danish fire insurance data (1980-1990) (in millions DKK). As originally proposed by McNeil we have fitted distributions on losses over 1 million of DKK.

Data are displayed in Figure 1 and some summary statistics are in Table 1. It could be noticed a significant skewness of the empirical distribution and only three claims larger than 65.7 million.

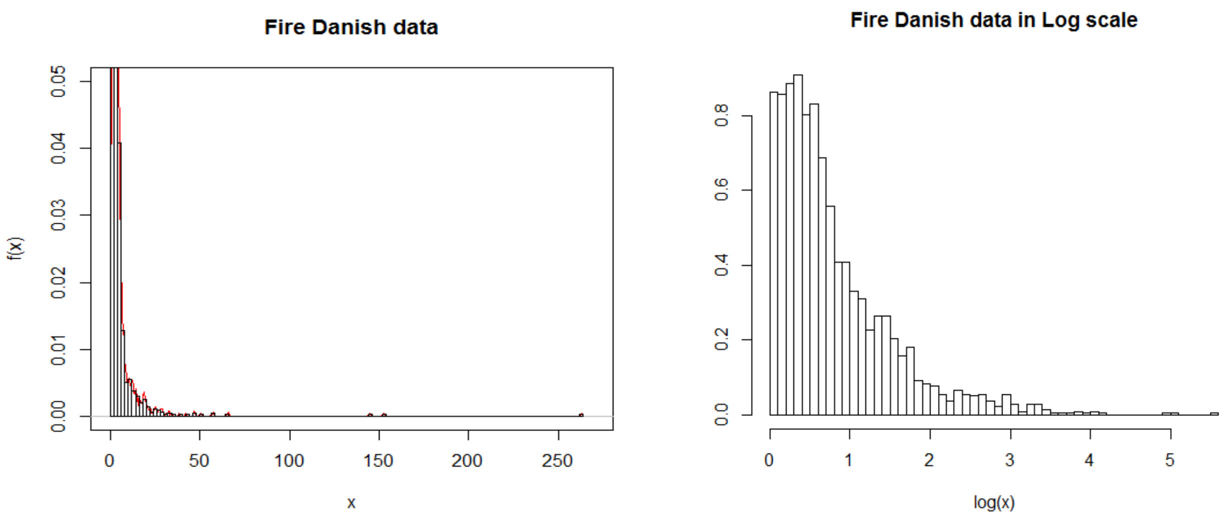

Figure 1. *Empirical distribution of Fire Danish Claims*

Num. 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

Table 1. *Main characteristics of empirical distribution*

Next figures compare classical distributions fitted to data by using maximum likelihood methodology and the basic minimum distance approach (1) based on Anderson-Darling distance function (i.e. in (2) $q = 2, p = 1$).

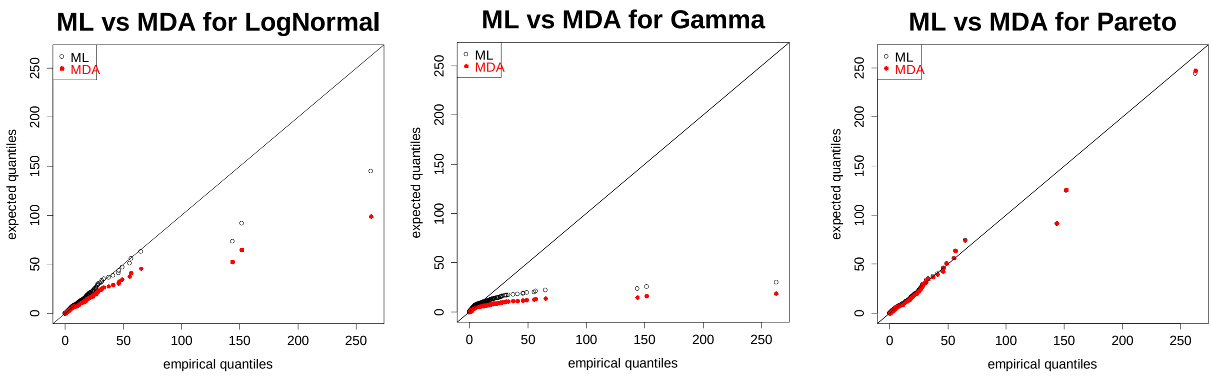

Figure 2. *MDA (AD) vs ML for different univariate distributions (Lognormal, Gamma and Pareto)*

Table 2 compares the fitting using D for ML and MDA and the corresponding estimated Akaike Information criteria, AIC = $AIC = 2\,\text{loglik} - 2k$, where k are the number of parameters..

	D for ML	D for MDA	AIC
LogNormal	149.4742	213.23.08	6732.918
Gamma	309.8396	339.1291	7428.887
Pareto	65.08656	64.35078	6683.403

Table 2. *Fitting measures for univariate distributions*

Both LogNormal and Gamma distribution fail to describe the shape and the tail of empirical values. However it could be noted that Pareto distribution fitted by using MDA tend to reproduce ML estimation and both reduce the underestimation on the tail.

To investigate whether the generalized distance (2) returns a distribution with a better fit, we have evaluated the behavior of loss function according to different combination of p. For the sake of simplicity, we picture in Fig.3 the loss D for q=1,2 and for different p for the only LogNormal distribution. Similar results have been derived for the other distributions.

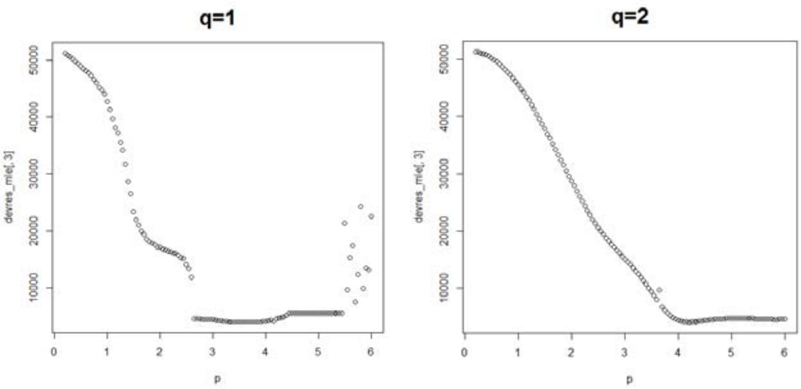

Figure 3. *D for different p and q equal to 1 or 2 for the lognormal distribution*

In both cases it emerges that a value of $p \cong 4$ assures the minimum of (2). This will lead to put a great weight on the tails.

Next figures compare classical distributions fitted to data by using maximum likelihood methodology and the minimum distance approach based on Anderson-Darling distance with p related to the minimum of D.

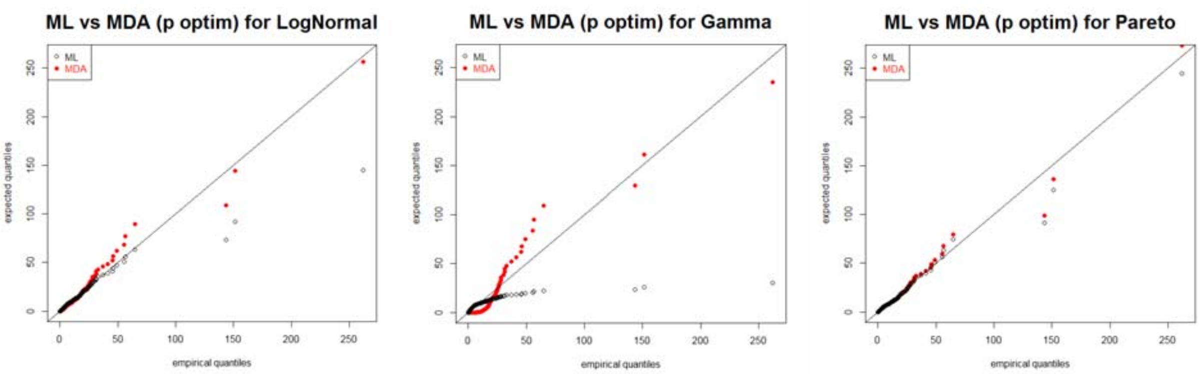

Figure 4. *MDA (AD) with optimum p vs ML for different univariate distributions (Lognormal, Gamma and Pareto)*

From table 3 an overall improvement of the fit is clear and form fig. 4 also the tails are better represented by using MDA for both LogNormal and Pareto

	D for ML	*D* for MDA	opt *p*
LogNormal	149.4742	63.55198	4.2
Gamma	309.8396	155.0078	4.35
Pareto	65.08656	55.47743	1.2

Table 3. *D for univariate distributions using the generalized distance (2)*

It is remarkable that in the case of Pareto the weight is around 1, which naturally leads to AD distance.

Considering mixtures of distributions, we report in Fig.5 the main results by using a LogNormal-Pareto mixture. The MDA approach seems to perform better than ML in fitting extremes values. Moderate (positive) bias exists in fitting attritional losses.

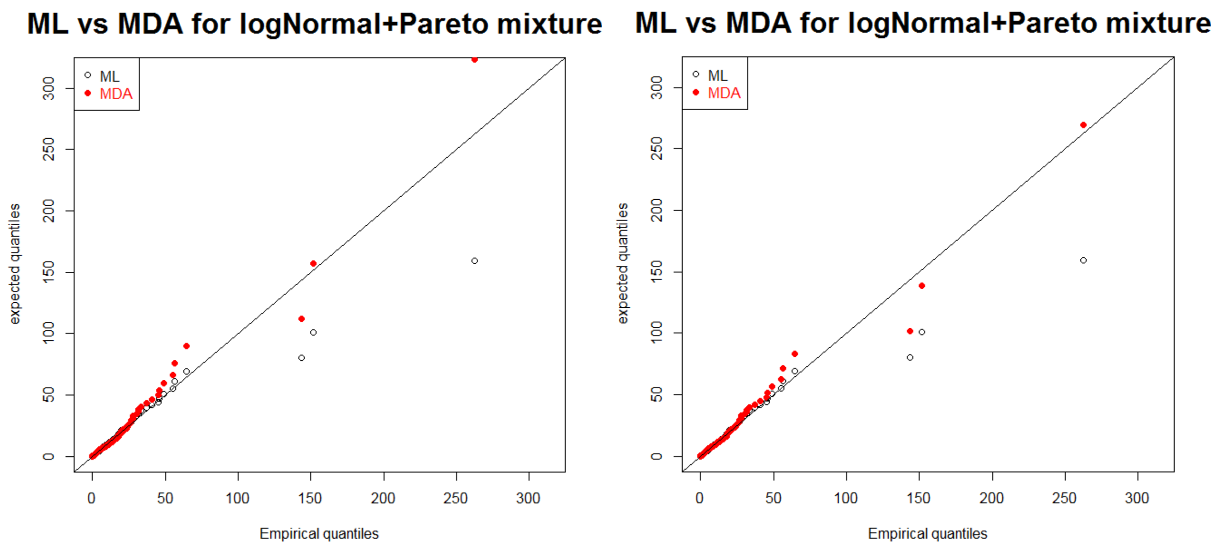

Figure 5. *MDA vs ML for the mixture Lognormal-Pareto (on the right the mixture for the generalize distance (2) with optimum p)*

ML	MDA (*p*=1)	MDA (*p*=3.4)
131.39697 (π=0.389)	80.51 (π=0.082)	54.14 (π=0.456)

Table 4. *Fitting measures (D) for the mixture LogNormal-Pareto (AIC=*3320.9870233*). In brackets the estimate of the mixing parameter.*

From Table 4 it emerges that the mixture with p=3.4 has increased further the fitting to the data by reducing the value of D in a significant way.

Analogously to previous comments a comparison based on a LogNormal-Pareto spliced distribution has been investigated. The D fitting measures in Table 5 are worst than the ones obtained with the mixture distribution. Nevertheless the most result is that the quantile corresponding to the π parameter may be interpreted as the threshold to discriminate between attritional and large claims, resulting in a procedure that natively helps in the selection of large losses. The main drawback is that at present the computational time is still too large and needs further research to obtain efficient and reasonably fast procedures.

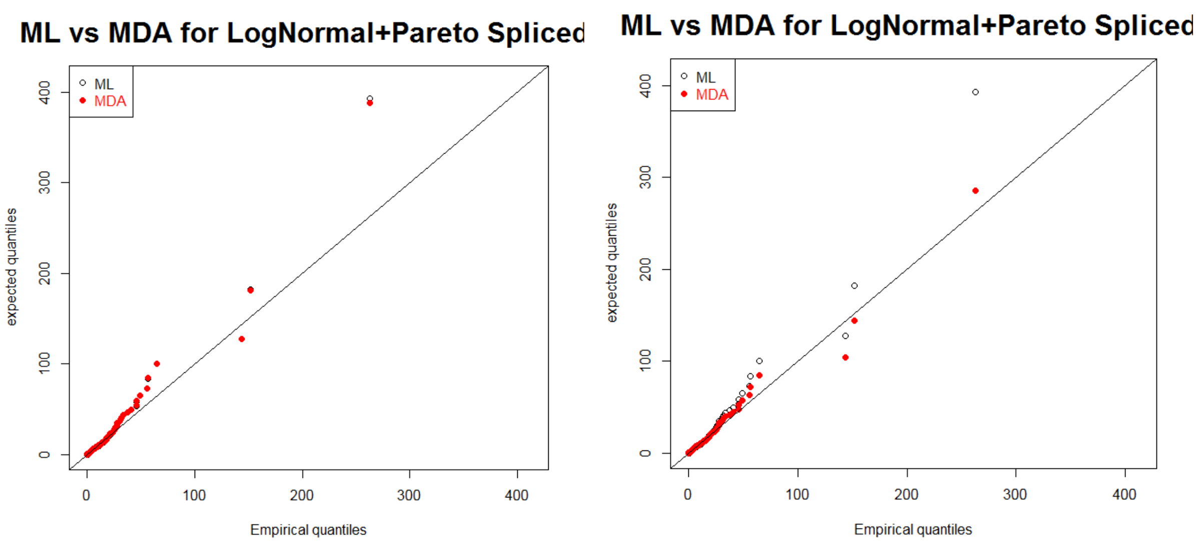

Figure 6. *MDA vs ML for the spliced Lognormal-Pareto (on the right the mixture for the generalize distance (2) with optimum p)*

ML	MDA (p=1)	MDA (p=3.6)
147.84 (π=0.48)	143.67 (π=0.544)	56.82 (π=0.709)

Table 4. *Fitting measures (D) for the mixture LogNormal-Pareto (AIC=*3316.162). *In brackets the estimate of the mixing parameter.*

Conclusions.

By using the well-known Danish fire claims data, we explored the capability of the Minimum Distance Approach to fit pure, mixtures and spliced distributions. 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).
Main results show that MDA could assure a better fit than ML due to its natural property to adapt the distribution to the data. MDA is data driven than it fits to the natural shape of the empirical distribution and moreover it is easy to implement, in some cases, even in a spreadsheet.
On the other hand, asymptotic distributions of estimators are not easy to be derived in closed form. Bootstrap procedures are often needed. Furthermore, if no weights are used and dataset is large, information from relevant but scarce data (e.g. extremes) are lost by the procedure.
In general 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.
Further development will regard both an analysis of variability of estimators and the evaluation of the effects of a different calibration of the severity distribution on the aggregate claim amount and hence on the capital requirement for premium risk.

References

[1] Basu A., Shioya H., Park C. (2011), *Statistical inference: the minimum distance approach,* CRC Press, 2011;

[2] Charpentier A., Oulidi A. (2010), *Beta Kernel quantile estimators of heavy-tailed loss distributions,*Statistics and Computing, 35-55

[3] Cooray K., Ananda M M A (2005), *Modelling actuarial data with a composite LogNormal-Pareto*, Scandinavian Actuarial Journal 5, 321-334;

[4] Cooray K. (2009), *The Weibull-Pareto Composite Family with Apllications to the analysis of Unimodel Failure Rate Data*, Communications in Statistics – Theory and Methods 38, 1901-1915

[5] Embrechts P., Kluppelberg C., Mikosh T. (1997), *Modelling extremal events for insurance and finance,* Springer, Berlin

[6] Frigessi A., Haug O, Rue H (2002), *A dynamic mixture model for unsupervised tail estimation without threshold selection*, Departement of Mathematical Science, Norway;

[7] Klugman S., Panjer H.H., Wilmot G.E. (2010), *Loss Models: From Data to Decisions*, John Wiley & Sons;

[8] Hewitt C., Lefkowitz B. (1979), *Methods for Fitting Distributions to Empirical Data,* Proceeding for the Casualty Actuarial Society, LXVI

[9] Luong A., Garrido J. (1993), *Minimum Quadratic Distance Estimation for a parametric family of discrete distributions defined recursively*, Australian Journal of Statistics 35(1), 59-67

[10] McNeil A (1997), *Estimating the Tails of Loss Severity Distributions using Extreme value Theory*, ASTIN Bullettin 27, 177-137;

[11] Nadarajah S, Bakar A A S (2012), *New composite models for the Danish-fire data,* Accepted and to appear on Scandinavian Actuarial Journal;

[12] Parr W.C., Shucany W. R. (1980), *Minimum Distance and Robust Estimation,* Journal of the American Statistical Association, Vol 75, n. 371, 616-624;

[13] Parr W.C.(1985), *Minimum Distance Estimation,* Encyclopedia of Statistical Science, 1985, 5.

[14] Pigeon M., Denuit M. (2011), *Composite LogNormal-Pareto model with random threshold*, Scandinavian Actuarial Journal 3, 177-192;

[15] Resnick S. (1997), *Discussion of the Danish data on Large Fire Insurance Losses,* ASTIN Bullettin 27, 139-151

[16] Scollnick D. P. M. (2007): *On a composite LogNormal-Pareto model*, Scandinavian Actuarial Journal 1, 20-33;

[17] Scollnick D. P. M., Sun C. (2012), *Modelling with Weibull-Pareto Models,* North American Actuarial Journal, Volume 16, 2, 260-272

[18] Wand, M. P., and Jones, M.C., *Kernel smoothing*, 1995, Chapman & Hall, London.