The Czeledin Distribution Function

Markus Knecht* and Stefan Küttel†

August 22, 2003

Abstract

In the insurance and reinsurance industry, Normal and LogNormal distributions are widely used to model loss ratios. This might lead to an underestimation of the tail. To address this problem we propose a new distribution which we call Czeledin distribution. It is a combination of a LogNormal and a Pareto distribution.

1 Introduction

In the insurance and reinsurance industry Normal and LogNormal distributions are widely used to model losses or loss ratios. This might lead to an underestimation of the tail, which means that scenarios with a high loss ratio have probabilities which are far too low. For example, using a LogNormal distribution with a mean of 80% and a standard deviation of 20% gives a probability of 0.372% for loss ratios larger than 150%. Using a Normal distribution with the same parameters leads to a even smaller probability of 0.023%.

We therefore want to construct a new loss distribution with the following properties:

- similar behavior around the mean as a Normal or LogNormal distribution, which does not shift mean and standard deviation too much
- heavy tail (tail index > 0)
- continuous with existing density

*Converium Ltd, General Guisan Quai 26, CH-8022 Zürich, Switzerland, e-mail: markus.knecht@converium.com

†Converium Ltd, General Guisan Quai 26, CH-8022 Zürich, Switzerland, e-mail: stefan.kuettel@converium.com

- easy to handle inverse distribution

We decided to use a LogNormal distribution up to a certain point and a Pareto distribution after this point. This distribution $\mathcal{C}[\mu, \sigma, \kappa]$ we call a *Czeledin* distribution. The parameters μ and σ are mean and standard deviation of the underlying LogNormal distribution, and κ is called the *crossing point*. The crossing point κ defines the point where the distribution changes from a LogNormal to a Pareto distribution.

2 Definitions

In mathematical terms, the CDF of the Czeledin distribution is given by

$$\mathcal{C}[\mu, \sigma, \kappa](x) = \begin{cases} \mathcal{L}[\mu, \sigma](x) & : \quad x < \kappa \\ \mathcal{L}[\mu, \sigma](\kappa) + [1 - \mathcal{L}[\mu, \sigma](\kappa)]\mathcal{P}[\alpha, \kappa](x) & : \quad x \geq \kappa \end{cases}$$

where $\mathcal{L}[\mu, \sigma]$ is a LogNormal distribution with mean μ and standard deviation σ, and $\mathcal{P}[\alpha, \kappa]$ is a Pareto distribution with tail index α and threshold κ,

$$\mathcal{P}[\alpha](x) = 1 - \left(\frac{\kappa}{x}\right)^{\alpha} \qquad \text{for } x > \kappa.$$

To ensure differentiability of the Czeledin distribution function, the tail index α must be chosen to be

$$\alpha = \frac{e^{-\frac{\mu_0^2}{2\sigma_0^2}}}{\sqrt{2\pi}\,\sigma_0\,\mathcal{N}\left(\frac{\mu_0}{\sigma_0}\right)}$$

with

$$\sigma_0^2 = \log\left(1 + \frac{\sigma^2}{\mu^2}\right)$$

$$\mu_0 = \log\frac{\mu}{\kappa} - \frac{\sigma_0^2}{2}$$

and $\mathcal{N}(x)$ is the Standard Normal distribution.

The density of $\mathcal{C}[\mu, \sigma, \kappa]$ is then given by

$$\mathcal{C}_d[\mu, \sigma, \kappa](x) = \begin{cases} \mathcal{L}_d[\mu, \sigma](x) & : \quad x < \kappa \\ (1 - \mathcal{L}[\mu, \sigma](\kappa))\mathcal{P}_d[\alpha](x) & : \quad x \geq \kappa \end{cases}$$

where $\mathcal{C}_d$, $\mathcal{L}_d$ and $\mathcal{P}_d$ are the densities of the relevant distributions.

The inverse distribution function is also useful, especially for simulation purposes. It is given by

$$\mathcal{C}^{-1}[\mu,\sigma,\kappa](p) = \begin{cases} \mathcal{L}^{-1}[\mu,\sigma](p) & : \quad p < p_0 \\ \mathcal{P}^{-1}[\alpha](\frac{p-p_0}{1-p_0}) & : \quad p \geq p_0 \end{cases}$$

with $p_0 = \mathcal{L}[\mu,\sigma](\kappa)$, and α as defined above.

The first and second moments of the Czeledin distribution $\mathcal{C} = \mathcal{C}[\mu,\sigma,\kappa]$ are given by

$$\begin{aligned} E[\mathcal{C}] &= \frac{\alpha \cdot \kappa}{\alpha - 1} \cdot \mathcal{N}\Big(\frac{\mu_0}{\sigma_0}\Big) + \mu \cdot \mathcal{N}\Big(-\frac{\mu_0 + \sigma_0^2}{\sigma_0}\Big) \\ E[\mathcal{C}^2] &= \kappa^2\, e^{2(\mu_0+\sigma_0^2)} \cdot \mathcal{N}\Big(-\frac{\mu_0 + 2\sigma_0^2}{\sigma_0}\Big) + \frac{\kappa^2\alpha}{\alpha - 2} \cdot \mathcal{N}\Big(\frac{\mu_0}{\sigma_0}\Big) \end{aligned}$$

Obviously the first moment exists only if $\alpha > 1$, and the second moment only if $\alpha > 2$.

3 Fitting

3.1 Maximum Likelihood Fit

Suppose $x_1 < \cdots < x_n$ is a set of observed values. Let $f(x,\mu,\sigma)$ be the density of the LogNormal distribution with parameters μ and σ and $g(x,\alpha)$ the density of the Pareto distribution with tail index α. As seen in Section 2, the tail index α of the Czeledin distribution is a function of μ, σ and κ. Hence the likelihood function of Czeledin distribution is

$$L(\mu,\sigma,\kappa) = \sum_{i=1}^{m(\kappa)} \log f(x_i,\mu,\sigma) \quad + \sum_{j=m(\kappa)+1}^{n} \log g(x_j,\alpha(\mu,\sigma,\kappa))$$

where the index $m(\kappa)$ is defined by the condition that $x_{m(\kappa)} \leq \kappa$ and $x_{m(\kappa)+1} > \kappa$.

As the number of addends of each sum in the likelihood function is depending on κ, there is no analytical solution to maximize L, but has to be done numerically.

In addition to the above mathematical problems, there are often practical problems for such a fit. In case one has only a small set of observed values it makes little sense to fit three parameters. Or, if there are no extreme events in the data set, the crossing point κ would be very high, and thus there is little difference between a Czeledin and a LogNormal distribution.

3.2 Practical Guidance to Select Parameters

We propose to proceed in two steps: First, fit μ and σ and second, determine κ.

To fit μ and σ one has to remove all extreme events from the data set (if there are any) and then fit a LogNormal distribution to the remaining data.

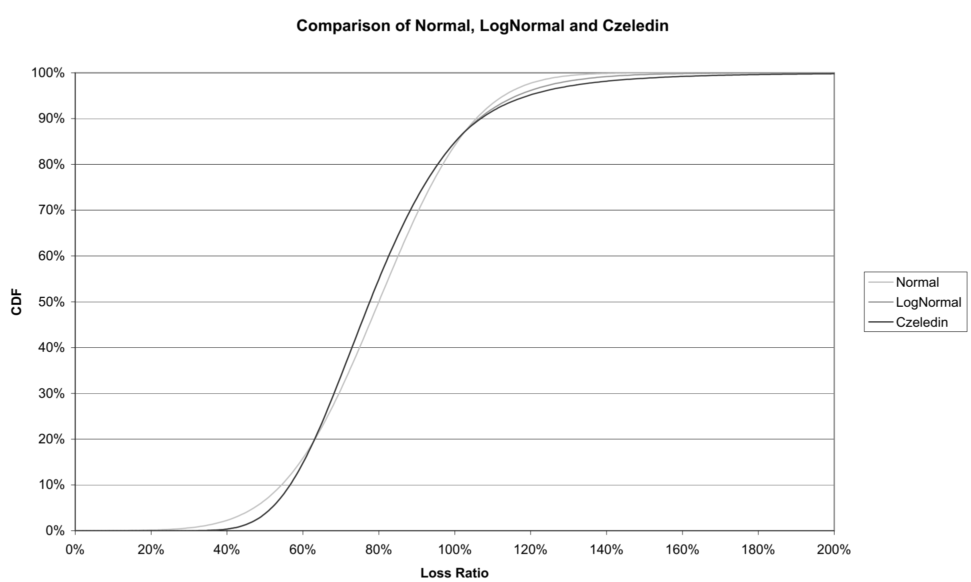
Now the extreme events are used to fit the tail index α and then to calculate κ from it. Unfortunately this can not be solved analytically, but has to be done using iterations or other numerical methods.

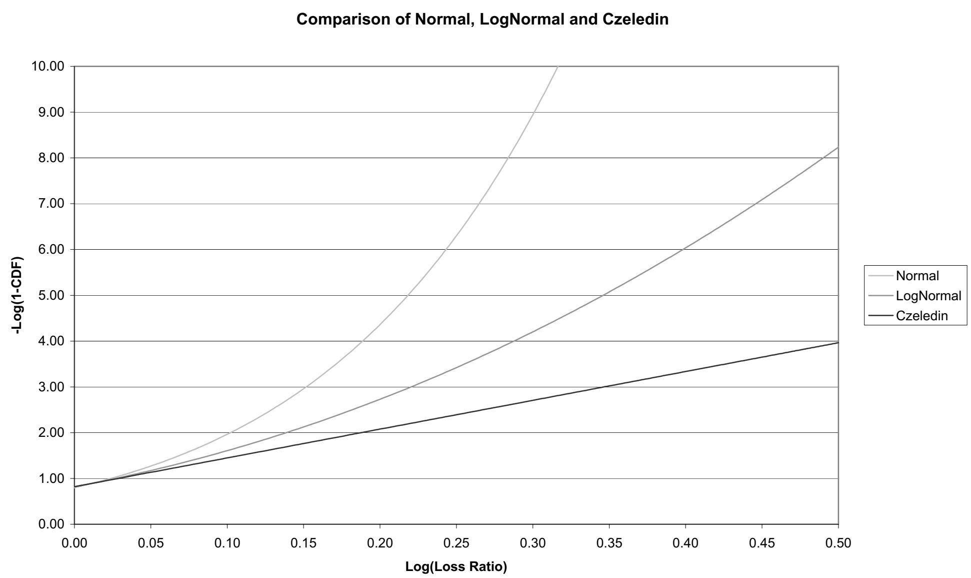
In case there are no or not enough extreme events there might be industry data to come up with a tail index α. Another approach is to select the crossing point a certain number of standard deviations above the mean: $\kappa = \mu + c \cdot \sigma$, with any positive parameter c. We found a good balance using $c = 1$.

4 Example

The following table and graph show a comparison of the Normal, LogNormal and Czeledin cumulative distribution functions with $\mu = 80\%$ and $\sigma = 20\%$. The crossing point was selected at $\kappa = \mu + \sigma = 100\%$. Notice: whereas for Normal and LogNormal distributions, μ and σ are the mean and standard deviation, this is not the case for Czeledin. For the above parameters, the Czeledin has mean of 80.69% and a standard deviation of 22.41%. The tail index is $\alpha = 6.290$.

Loss Ratio	Normal	LogNormal	Czeledin
80%	50.00%	54.90%	54.90%
90%	69.15%	72.62%	72.62%
100%	84.13%	84.84%	84.84%
110%	93.32%	92.17%	91.67%
120%	97.72%	96.16%	95.18%
130%	99.38%	98.19%	97.09%
140%	99.87%	99.17%	98.17%
150%	99.98%	99.63%	98.82%
160%	100.00%	99.83%	99.21%
170%	100.00%	99.93%	99.46%
180%	100.00%	99.97%	99.62%
190%	100.00%	99.99%	99.73%
200%	100.00%	99.99%	99.81%
250%	100.00%	100.00%	99.95%
300%	100.00%	100.00%	99.98%
400%	100.00%	100.00%	100.00%





5 Summary

LogNormal distributions usually fit well for attritional loss distributions so that, ideally, large losses should be modelled separately.

If this is not possible, and as large losses tend to make the tail of the distribution heavier, the proposed Czeledin distribution has in our experience proved to be a good approach for modelling a heavier tail.

The example in the Section 4 shows that a typical Czeledin distribution gives reasonable probabilities even for scenarios with loss ratios of 200% or more. Such scenarios are regularly observed in the insurance and reinsurance industry.