How to validate op risk distributions

Polling experts on duration works best for assessing operational risk problems because it fits neatly with the way non-statisticians think. By Carsten Steinhoff and Rainer Baule

Carsten Steinhoff,
Nord LB

Rainer Baule,
University of Goettingen

OPERATIONAL risk is on the rise. For a long time, this type of risk was neither explicitly regarded nor measured, even though it is one of the oldest risk types a bank faces. Due to spectacular losses such as those faced by Barings or recent regulatory developments in Basel II, a new focus has been put on op risk, and methods for its treatment are growing quickly.

The quantification of op risk is an especially difficult task. Due to the inadequacy of most internal data sources in covering extreme losses, it is essential to consider not only real-life data (from internal or external sources), but also synthetic data (obtained from scenario analysis).[1] This data, of different types and quality, has to be properly mixed to get a representative dataset.

In a second step, we try to match this dataset with a parametric distribution. Within loss distribution approaches, usually one sub-model is chosen for the distribution of the severity of losses and one for the frequency of losses. While the distributional frequency analysis is commonly regarded as easy, it is tricky to find appropriate severity models that cover extreme events adequately. Many publications written by practitioners, as well as by academics, suggest the use of extreme value distributions for the tail-modelling problem.

All utilised distributions have to be tested for their appropriate correspondence to reality. This is often done by experts or by a cross-check with data from other sources. An 'expert' within this scope is an employee of the bank who has particular expertise about a certain type of business and the associated risks, but is not necessarily familiar with statistical methods.

A considerable problem thereby is how to compare subjective impact from expert assessments and (more or less) objective distribution assumptions. By their nature, both views are quite different. On the one hand, there is the pure statistical view of a (parametrical) distribution function. On the other hand, there is the expert's practical intuition, expressed in opinions about how often a risky event might occur and how severe it might be.

The question addressed in this article is: can both views be synchronised? Surely, the answer depends on the objective. If only the identity of two distributional assumptions needs to be compared, tools like QQ-plots or step-plots are very helpful.[2] But they will not help the experts to understand why and how an objectively parameterised (distributional) severity model should be adjusted to their opinion.

Our aim is to present a simple representation of distributional assumptions that can be easily understood and modified by experts. Hence, expert opinions and distributional assumptions can be cross-checked. Furthermore, parameter fittings can be represented transparently to non-statisticians.

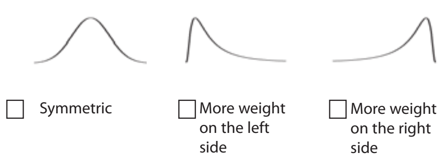
Representing subjective opinions

If we were to directly ask an expert which distribution is appropriate to a certain risk category, we would normally receive no useful answer. For instance, choosing between some distributional forms as shown in figure 1 would coerce them to think in statistical dimensions that are unfamiliar to them.

Although this form of questioning is, as far as we know, not very popular in the banking industry, it is a good example of a representation that is highly correct from the model perspective, but incomprehensible for 'normal' experts. Another line of questioning that is more often used is to ask experts how many events could happen in different risk classes (see table A).

1. An expert's choice on the distribution shape

Question: Which representation best fits the risk in your department?



[1] This is also required by Basel II. See BCBS (2004)
[2] For a good introduction to quantification topics and practical problems within the operational risk context, see Frachot *et al* (2003)

A. Elicitation based on grouped data

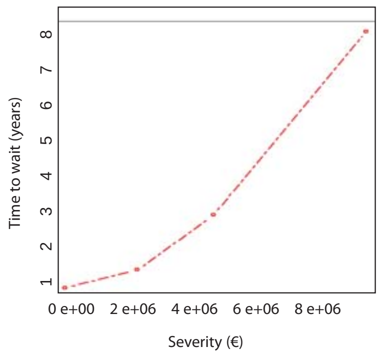
Question: How many risky events happen per year in the following classes? Which amount of loss will be most likely?

From	To	Most likely severity	Expected frequency		
			Min	Ave	Max
€1,000	€100,000				
€100,000	€1 million				
€1 million	€10 million				
€10 million	Maximum				

B. Example for an elicitation with duration

Severity exceeded	Duration of event
€100,000	1
€2 million	1.5
€4 million	3
€8 million	8

2. Graphical representation of table B



Usually, an expert remembers more or less how often an event occurred in the past few years or can at least express an opinion on the supposed occurrence. Hence the risk analyst can transfer the opinions obtained by the expert poll into a distribution (parametric or non-parametric). This task will not be described here. Nevertheless, we think it is worthwhile to reflect on the cognitive process that an expert would utilise to fill in the table in table A. Because the table asks for absolute quantities, the expert would have to solve a multi-stage cognitive problem. Intuitively, we could describe the questions in the process as follows[3]:

1. Which events of severity between €x and €y do I remember?
2. How many of those events have happened in a year on average?
3. How many of those events happened in a good year (minimum) and in a bad year (maximum)?

These questions are quite easy if, on average, there is more than one event a year. However, the crucial losses are usually the low-frequency, high-impact events that occur very rarely. For those events, an additional step has to be introduced to the process to elicit how many events occur in one year:

2a. How many events happened in the past 10 or xx years?
2b. Divide this figure by the number of years.
3a. To obtain maximum and minimum figures, step 3 has to be altered to: how many events can happen in the chosen time range in worst- and best-case scenarios?

According to our experience most experts 'fail' in step 2a. They do not ask themselves this question rigorously, but rather try to answer it intuitively. And if they do, they process step 3a by adjusting the average with a sense of proportion to get a minimum or maximum.

As mentioned, in operational risk modelling especially, an evaluation of extremely severe losses is focused. These events occur very rarely, with a frequency of lower than one event a year. Considering that accuracy in step 2a and 3a is highly relevant to obtain credible results, which is problematic as discussed above, the methodology shown in table A is more appropriate for evaluating medium-sized losses than it is for large ones.

We are confident that a method that directly asks for the duration of one event should lead to more precise results. We define the duration as mean time to the occurrence of one event exceeding a certain severity. The question to ask is now: how many years will we have to wait, all things being equal, to observe an event of severity x (or above)? For example, we may obtain table B from an expert assessment.

This can easily be represented graphically. Figure 2 directly shows what the expert thinks – for example, that a loss of €4 million happens once in three years.

Comparison of parametrical model and expert opinions

Making representations comparable

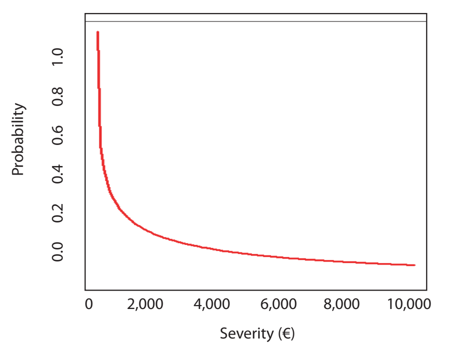
The aim of this section is to combine the expert's opinion and the parameterised model. We not only want to compare the two sources by means of a cross-check, but also want to strive to give the expert a chance to adjust his initial estimates directly. This second step should lead to a new (parametric) estimate for the severity distribution.

Therefore, we have to transform the distributional representation into a form the expert understands easily.

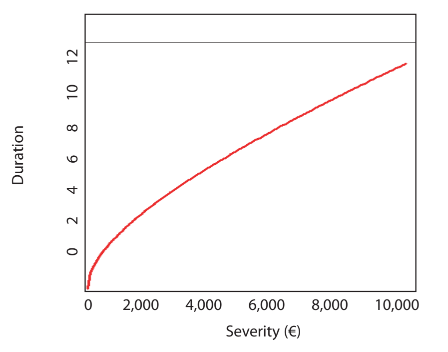
It is quite easy for a given distribution to generate a corresponding duration table similar to table B or its graphical representation as shown in figure 2. This seems to be straightforward, but as far as we know this approach is not very widespread in the banking industry.

[3] Because we want to focus on the representation of distributions to experts we will not describe further psychological aspects (eg, the impact of observed events to the estimate, etc). As a good reference, see Kahneman *et al* (1982)

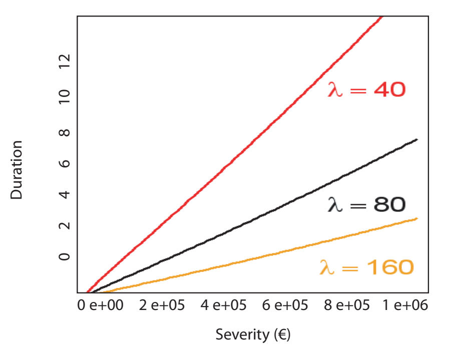
3. Excess function of a lognormal distribution (μ=5, σ=3)



4. Duration for one event in a log-normal distribution (μ = 5, σ = 3)[4]



5. Duration for different frequencies



Any parametric distribution $F(x;\Theta)$ can be transformed into a so-called excess function. This depiction provides the information with which probability a certain threshold is exceeded:

$$P(X \geq x) = 1 - F(x;\Theta)$$

An example for a lognormal distribution with parameters μ = 5 and σ = 3 yields is shown in figure 3 (To simplify the matter for demonstration we supose that exactly one loss will happen a year).

To get a representation equal to the expert's opinion, the figure's y-axis has to be transformed from a probability into a frequency scale, which is simply the reciprocal of the previous equation (see figure 4):

$$d_1(x) = \frac{1}{1 - F(x;\Theta)}$$

So far, we have only considered the severity distribution. Hence the calculation gives us the average time d_1 we have to wait until an event of severity x occurs if the annual event frequency equals 1. If this frequency is Poisson distributed with parameter λ, there are λ events per year on average, hence the duration decreases by a factor λ:

$$d(x) = \frac{1}{\lambda\left(1 - F(x;\Theta)\right)}$$

Figure 5 shows this duration for different values of λ. The higher λ is, the less steep the duration function will be. The idea behind this plot is identical to the treatment of the expert poll in table B. The expert can now decide whether he wants to adjust his estimate or not. The curves are also easily comprehensible for non-statisticians and could be communicated very well to board members, for example. With this representation the model process will become more transparent than it could ever be with 'traditional' distribution plots. At the same time, Basel's demand for a transparent model is accomplished.

Fitting various distributions

Beyond a simple visualisation, opinions can also be transformed directly into parametrical distributions. That opens two possibilities for the evaluation. First, as shown above, the expert's parameters can be compared with the 'objective' parameters, giving the expert the opportunity to judge how reality is described best.

Second, a distribution can be derived directly either from only the expert's estimates or from a 'virtual' database that mixes various sources.

In the first case, the aim is to check the expert's estimates with respect to a more objective source. The second case focuses on the accuracy of the mathematical fitting. For the fitting, for example, an optimisation approach can be used, which minimises an error functional based on theoretical and empirical durations d_i. This error functional could, for instance, measure the mean squared error:[5]

$$\Theta = \arg\min_{\Theta} \sum_i \left(d_i - \frac{1}{\lambda\left(1 - F(x_i;\Theta)\right)} \right)^2$$

Various distributions can be checked and compared analytically as well as graphically.

[4] If we assume an average annual frequency of one event per year, the minimum possible duration is exactly one
[5] An alternative to this absolute mean squared error could be the relative mean squared error

Example

The following short but realistic example sketches the idea of our approach. One of our experts has estimated the figures in table C.

This estimation can be fitted with good accuracy by a generalised Pareto distribution with parameters $\xi = 0.8$ and $\beta = 500{,}000$.[6]

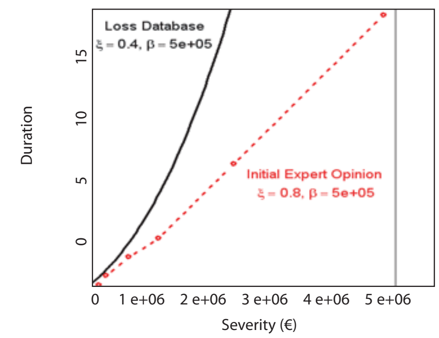
According to our analysis, the historical database of this example is fitted by the same distribution with $\xi = 0.4$ and $\beta = 500{,}000$. A comparison of both views gives figure 6.

The expert can now decide whether he believes more in his own estimate d_{EXP} or the parametrical representation d_{DB} derived from the database. For example, he could judge that the adjusted estimate lies exactly between both representations. In this case, the 'new' duration would be given by $0.5\ d_{EXP} + 0.5\ d_{DB}$, if we assume a linear combination.[7] With this point of view it is also possible to let the expert adjust his estimate to different scenarios, for example the three generical cases: normal, adverse and crucial/extreme.

C. Example for an elicitation with duration

Severity of one event	Happens once every... years
€375,000	1.5
€500,000	2.0
€875,000	3.0
€1,375,000	4.0
€2,625,000	8.0
€5,125,000	16.0

6. Confronting the expert's opinion with a database



Summary

This short article on the representation of distributions in an operational risk context gives a very basic overview of how risk managers can obtain estimates from experts.

We have shown different sorts of expert polls and found that the duration is best for operational risk problems because it fits best to the way a non-statistician thinks and thus provides a better accuracy for the estimate of rare events.

We stated that distribution functions of any type can be represented as exceedances, too, and thereby are highly comparable with the experts' estimates. On the one hand, experts can directly compare their own estimates to a distributional assumption. That way, the modelling task is no longer solely a quant's playground but can also be partly done by people who are very close to a certain risk.

On the other hand, risk managers can analyse the fitting and credibility of curves very easily and depict their results in a way that is very comprehensible for non-statisticians as well.

[6] See Embrechts *et al* (2003) for the parametrisation of this distribution
[7] That approach is very close to the so-called 'credibility theory', which is very popular in actuarial science. The mixing is also possible if expert and database distributions are different by type

Carsten Steinhoff is op risk analyst at NORD/LB. Email: carsten.steinhoff@nordlb.de. Rainer Baule is a postdoctoral research fellow at the University of Goettingen and managing director of Deriva Financial IT and Consulting. Email: rbaule@wiso.uni-goettingen.de. The views expressed here are those of the authors and do not necessarily reflect those of the NORD/LB. All examples are fictitious.

References and further reading

BCBS, 2004
International convergence of capital measurement and capital standards: a revised framework
Basel

Chavez-Demoulin V and P Embrechts, 2004
Advanced extremal models for operational risk
Zürich

Community of European Banking Supervisors, 2006 *Guidelines on the implementation, validation and assessment of Advanced Measurement (AMA) and Internal Ratings Based (IRB) approaches*
Consultation Paper 10 as amended on January 20, 2006, London

Cruz M, 2002
Modeling, measuring and hedging operational risk
West Sussex

Embrechts P, H Furrer and R Kaufmann, 2003
Quantifying regulatory capital for operational risk
Working paper, Zürich

Frachot A, O Moudoulaud and T Roncalli, 2003
Loss distribution approach in practice
Paris

Kahneman D, P Slovic and A Tveryky, 1982
Judgement under uncertainty
Cambridge