Fitting Distribution Scenario Analysis Interval Approach

08 NOV 2011 by Dinesh Chaudhary

Pristine (www.edupristine.com) and Bionic Turtle have entered into a partnership to promote practical applications of the concepts related to risk analytics and modelling. Practical and hands on understanding of building excel based models related to operational and credit risk is necessary for any job related to risk management. For this purpose, we would be illustrating step by step model building techniques for risk management. Registrations for Operational Risk are OPEN.

Operational Risk Modelling Analytics
Fitting distribution to Scenario Analysis Data in Excel – I
Using Interval Approach

Hi,
This is Dinesh from Pristine! I will discuss how to fit distribution to data collected from operational risk experts through Scenario Analysis.

Introduction

Operational Risk VaR estimation under Basel–II Advanced Measurement Approach (AMA) requires data (Across eight business lines (BL) and seven event types (ET)) on

- Frequency and
- Severity of operational losses

Challenges

In practice, empirical loss data is missing/scant for many cells (a cells is a unique combination of BL and ET) to arrive at reliable distribution parameter estimates. Even when empirical data is present, it might need to be supplemented with forward-looking approach.

Scenario Analysis

Scenario analysis helps banks to gather data from business/subject matter experts about their estimate of frequency and severity of plausible operational risk events. It is especially applicable for Low Frequency High Severity events.

To be able to fit distribution to scenario data, it is important that data is elicited from experts in such a manner that it is amenable to distribution fitting using one of the methods such as moment matching, Quantile/percentile matching, maximum likelihood, OLS.
Business experts may not understand probability and statistics. So questions need to be framed in such a manner that scenarios are easy to understand and probability distributions can be fitted to frequency and severity data elicited from experts.

**Distribution fitting to scenario data**

There are usually three methods of collecting scenario data:

- **Interval approach**: experts mention the frequency of losses estimated within specific loss intervals
- **Percentile approach**: data is collected for specific percentiles/quantiles of loss severity from experts
- **Individual scenario approach**: individual data point linked with a frequency of occurrence

In this tutorial we would discuss the Interval Approach. In the following illustrations, we will fit continuous distributions to scenario data collected for loss severities.

**Using MLE for fitting severity distribution to data collected through interval approach**

Following are the consensus estimate elicited from business experts through a scenario analysis workshop for ET ‘Internal Fraud’ in BL ‘Commercial Banking’. Each cell represents estimated number of events in a loss interval. Experts may also be asked about the maximum impact of a single event.

<table>
<thead>
<tr>
<th>Loss Severity</th>
<th>0-7500</th>
<th>7501-1750</th>
<th>1750-3250</th>
<th>3250-6750</th>
<th>6750-1250</th>
<th>1250-3000</th>
<th>&gt;300000</th>
<th>Max. Single Event Loss (Million)</th>
<th>Total Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expected number of loss events</td>
<td>99</td>
<td>42</td>
<td>29</td>
<td>28</td>
<td>17</td>
<td>9</td>
<td>3</td>
<td>5</td>
<td>227</td>
</tr>
</tbody>
</table>

**Algorithm**

**Step-1**: Decide on a distribution to be fitted to data. For this illustration, let us fit a lognormal distribution to scenario data.

**Step-2**: Decide on seed values of distribution parameters to initiate likelihood function optimization. Refer [article 1](#) for an introduction to using MLE for distribution fitting. For lognormal distribution, both parameters need to be positive.

**Step-3**: Calculate the likelihood of observing severity of an event in the above severity bins/intervals as:

- Likelihood of observing loss in an interval = \( \text{CDF(Upper bound)} - \text{CDF(Lower bound)} \)
- Likelihood of observing 'N' data points in an interval = \( (\text{CDF(U)} - \text{CDF(L)})^N \)
- Taking log on both sides, log–likelihood of observing n data points in an interval = \( N \times \log(\text{CDF(U)} - \text{CDF(L)}) \)
Assuming operational risk loss events are independent, joint likelihood of observing all data points together is the product of individual likelihood. Taking logs simplifies the product function to a summation function. Therefore, joint log-likelihood function is the sum of individual likelihood function. **Step 4:** Use an optimization algorithm (like Excel Solver) to maximize joint log-likelihood by changing parameter values. Instead of maximizing log-likelihood, negative log-likelihood may also be minimized. Screen below shows the optimized value. Joint negative log-likelihood declines from 461.69 to 313.78, if parameter–1 (log_mean) is changed to 9.12 and parameter–2 (log_stdev) is changed to 1.36. If maximum loss is not elicited from the subject matter expert, then also it is easy to adjust the likelihood function. Likelihood on a single event’s severity to fall in the ‘>300000’ interval would be \(1 - CDF(300000)\).
Step-5: Scenario modeller may fit various continuous distributions to the interval data and the distribution with the least value of joint negative log–likelihood maybe taken as the best fit.

Templates to download

I have created a template for you, where the subheadings are given and you have to link the model to get the cash numbers! You can download the same from here. You can go through the case and fill in the yellow boxes. I also recommend that you try to create this structure on your own (so that you get a hang of what information is to be recorded).

Also you can download this filled template and check, if the information you recorded, matches mine or not!

Way forward

This tutorial illustrated the Interval approach to fitting distribution using scenario analysis. I hope that you enjoyed the tutorial. In the next tutorial we would illustrate the percentile approach. We would model the approaches in Excel and also discuss the challenges faced in the approaches. Stay Tuned.

By the way, we have launched the course on Operational Risk Modeling, which covers all these practical concepts through video lectures. If you want to join the same, you can join by clicking here.

To reach the author, you can send an email to dinesh.chaudhary@asymmetrix.co.in.

The case has been drafted for discussion purpose. It has been written by Pristine (www.edupristine.com) and would be discussed by experts from Pristine & Bionic Turtle. There would be a step by step analysis and financial model building to come to a conclusion on the decision.
Pristine is an authorized training provider for reputed organizations like CFA Institute (USA), PRMIA (USA), CARP (USA) and has provided trainings for reputed organizations like HSBC, Bank of America, JP Morgan, NUS, IIM Calcutta, etc.