A Simple Approach to Combining Internal and External Operational Loss Data

Pavel Okunev
Lawrence Berkeley National Laboratory,* UC Berkeley and Bank of America†

Abstract

We propose a simple approach to combining internal and external loss data in the case when internal and external data come from the same distribution. We assume that the internal data is uncensored but the external data includes only losses above a known threshold. This approach is an alternative to the method of Baud et al. [1], when the latter is too computationally expensive due to the large quantity of data available.

1. Introduction

In order to estimate the value at risk and the economic capital required under the Basel accords, banks need to model the distribution of their operational losses. This is achieved by finding a distribution from a given family or set of families which best fits a bank's historical loss data. However, internal data is often insufficient to accurately estimate the upper tail of the loss distribution, because extreme losses occur very rarely. The quality of approximation to the true loss distribution can be improved by incorporating external loss data. Such data is often available from a consortium of banks organized to collect and share loss data. It is usually assumed that internal and external data are sampled from the same distribution. It is additionally assumed that the internal data includes all the losses that occurred, while the external data includes only the losses exceeding a known consortium reporting threshold. It then becomes necessary to properly incorporate the internal and external data together.

1.1. Previous Work

A naive approach is to simply construct a new loss dataset containing all of the external and internal loss observations. However, this will tend to overestimate the likelihood of high losses because the external data are censored in a way which shifts the distribution of the external data to the right. Baud et al. [1] proposed a solution to this problem based on the maximum likelihood approach. While their approach eliminates the bias, it is too expensive computationally when the number of available loss data points is large.

2 The Stratified Sampling and Weighted Average Approaches

Here we propose two alternative approaches for solving this problem.

Example. Suppose we have 100,000 internal loss observations and 150,000 external observations drawn from the same loss distribution. Suppose the external observations have been censored and only the losses over $25,000 were included. The internal data has 50,000 losses less than $25,000 and 50,000 losses greater than $25,000. How can we use the internal and external data to obtain a sample from the loss distribution incorporating all of the available loss data?

This situation often arises in statistics in the context of stratified sampling [4]. There is a canonical solution to this problem. Combining internal and external data we obtain four times as many samples of losses over $25,000 as we had from internal data only. Therefore a sample from the loss distribution must contain all of the data for losses over $25,000 from the internal and external data and four copies of each data point for loss below $25,000 from the internal data. This new sample is not biased toward higher losses and incorporates all of the available information.

There is an alternative approach to estimating the loss distribution: the weighted average approach. The idea is that in computing empirical moments and quantiles of the loss distribution four times as much weight should be given to any loss under $25,000 from the internal data as to a loss over $25,000 dollars from either the internal or the external data.

The weighted average approach extends more naturally to the case when that number of external samples above the reporting threshold is not an integer multiple of the number of internal samples above the threshold.

*This work was supported by the Director, Office of Science, Office of Advanced Scientific Computing Research, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

†This work expresses opinions of its author and does not represent official views of Bank of America

3. Example and Conclusions

To test the proposed approach we do the following computational experiment. We draw 100,000 internal loss samples from the log-normal distribution with parameters $\mu = \ln(10,000) \approx 9.9034$ and $\sigma^2 = \ln(2) \approx 0.6931$. We also draw 25,000 external loss samples from the same distribution and censor the samples so that only losses over $25,000 are reported in the external loss data. This gives us 39618 internal loss samples with loss over $25,000, 9788 external loss samples all over $25,000 and 60382 internal loss samples with loss under $25,000.

We estimate the parameters of the log-normal distribution by computing the empirical mean and variance of the natural logarithm of losses. In this computation we give each internal loss under $25,000

$$1 + \frac{9788}{39618} \approx 1.2470 \tag{1}$$

times more weight than to any of the internal or external losses over $25,000. We arrive at the following estimate:

$$\mu = 9.9071 \tag{2}$$

and

$$\sigma^2 = 0.6950 \tag{3}$$

The results are very close to the true values of the parameters. This computation is considerably faster than the method of Baud et al. [1], when a large amount of data is available. The stratified sampling and the weighted average approaches can be valuable, when there is too much data to make the maximum likelihood approach of Baud et al. [1] practical.

4 Acknowledgments

I thank my adviser A. Chorin for his help and guidance during my time in UC Berkeley. I thank Mathilda Regan for help in preparing this article. I am also grateful to Mark Piche and John Walter for encouraging my interest in finance. Last, but not least, I thank my family for their constant support.

References

[1] N. Baud, A. Frachot, T. Roncalli, “Internal Data, External Data and Consortium Data for Operational Risk Measurement: How to Pool Data Properly? ” *Credit Lyonnais*, http://www.gloriamundi.org/

[2] N. Baud, A. Frachot, T. Roncalli, “How to Avoid Over-Estimating Capital Charge for Operational Risk? ” *Credit Lyonnais*, http://www.gloriamundi.org/

[3] N. Baud, A. Frachot, T. Roncalli, “An Internal Model for Operational Risk Computation ” *Credit Lyonnais*, http://www.gloriamundi.org/

[4] S. Thompson, Sampling, Wiley-Interscience , 1992.