A Note on Scenario Analysis in the Measurement of Operational Risk Capital: A Change of Measure Approach

David F. Babbel[1]

September 24, 2010

Abstract

Since I circulated a working paper (coauthored with Kabir Dutta) entitled "Scenario Analysis in the Measurement of Operational Risk Capital: A Change of Measure Approach" on the Wharton Financials Institutions Center website in March of 2010, many different questions have been received concerning the methodology we used. While we addressed those questions at an individual level, in this note I would like to address the questions for the readers at large.[2]

Key Words: Scenario Analysis, Operational Risk Capital, Change of Measure, Internal Loss data Modeling, Operational Risk Measurement

[1] David F. Babbel is a Fellow of the Wharton Financial Institutions Center, Professor at The Wharton School of the University of Pennsylvania, and Senior Advisor to Charles River Associates.
[2] I am grateful to Ravi Reddy and David Schramm for their valuable assistance.

The methodology introduced in "Scenario Analysis in the Measurement of Operational Risk Capital: A Change of Measure Approach" (hereafter, "our study")[3] has a number of vital components:

- Scenario data
- The methodology and its justification
- An evaluation of scenario data and its impact

In the following sections, I will discuss these components in order to address some very good questions we have been asked. In the process of elaboration, I cite some documents and use examples that are not in our study to illustrate the concepts we developed in our study.

The methodology of our study is based on a few very simple but ***key insights*** we introduced, which are:[4]

1. If, for example, a scenario is once in 20-year event and the historical data were collected for the last 10 years, one cannot simply merge the severity of the scenario data with the historical data. This will increase the probability of the severity of the scenario data by two fold. This goes the other way too. If a scenario is a once in 5-year event, by merging it with 10 years of historical data, the probability of the scenario will be diluted by half. Therefore, the duration of the scenario data should be properly normalized with the duration of the historical data. This is the key contribution of our study.[5]

2. It follows from (1) that if the frequency is distributed as Poisson with parameter λ, then a scenario with the chance of happening once in N years is $1 \div (N \times \lambda)$. Later, I will clarify this with an example.

3. The historical forecast for an event (a subset of R^+) can be adjusted to match with the corresponding forecast given in the scenario by reweighting the historical probability for every event, which in the literature is known as the Change of Measure approach.

Scenario Data

From sections 1.1, 1.2, 2, and 3 of our study:

In our approach scenario data are used to reweight the probability of tail loss events. This is consistent with the best practices for operational risk management and Basel regulatory require-

[3] We have made several updates to our study. Readers should access the latest version of the working paper available on the *Wharton Financial Institutions Center* website and also on the *SSRN* website.
[4] To the best of my knowledge, no published document has interpreted scenarios related to operational risk in this manner before our study.
[5] The concept and methodology given in the working paper are copyrighted but freely available for use with proper citation. Any improper use will be a violation of the copyright law.

ments. Therefore, it should be noted that in our approach a scenario has the following characteristics:

- A scenario data point is defined as a high severity and low probability (higher quantile of the severity distribution or tail) loss data point.
- A tail loss data point is typically at the 99th percentile or higher level of the current loss experience.
- By virtue of being tail events, scenario data points are very few in number in comparison to the number of internal loss data events experienced by an institution.
- Scenario data will contain either none or very few low severity and high probability data points (losses commonly found at the lower quantiles of the severity distribution).

For our approach we need the severity of the data represented in an interval in any of the following formats:

(a,b), (a,b], [a,b), [a,b], (a,∞), or [a,∞). While data in any of these formats can be used in our model, we would advise that data be collected in any of the following formats only: (a,b), (a,b], [a,b), or [a,b]. We have evidence that data collected in the format (a,∞) or [a,∞) are subject to several inconsistencies. Moreover, it is unrealistic to think that for many of the possible losses, the loss, at least in financial terms, can be unbounded.

Scenario analysis can be used for what-if analysis, stress testing, and, when appropriate, for forecasting. Institutions should decide how a scenario can best be used. It has been shown in research such as Kahneman, Slovic and Tversky (1982) that human prediction capability is at its best when thought of as within a bounded interval. Based on our own experiences at scenario generation workshops performed at several financial institutions, we find that the assertions in Kahneman, Slovic and Tversky (1982) also hold true in the context of operational risk scenario development.

The probability of a particular scenario occurring is often expressed in terms of the "number of years" until such an event occurs, such as once in twenty years (1 in 20). In our approach we have interpreted it as follows:

Suppose the average annual frequency of losses from the internal (*i.e.*, institution's) loss experience is 25 losses per year. Using the frequency interpretation of probability, we interpret the probability of this event happening as 1 in 500 (= 20 x 25) on average. In other words, we expect to see one such loss , on an average, for every 500 losses. One should note that it is for an average interpretation based on the assumption that loss frequency data are distributed as a Poisson distribution (in this case with $\lambda = 25$). In general terms, if an event happens once in N years and we assume that the frequency of the loss distribution is distributed as the Poisson distribution with parameter λ, then using the frequency interpretation of the probability we say that the probability of the event happening is, on average:

$$1 \div (N \times \lambda) \tag{1}$$

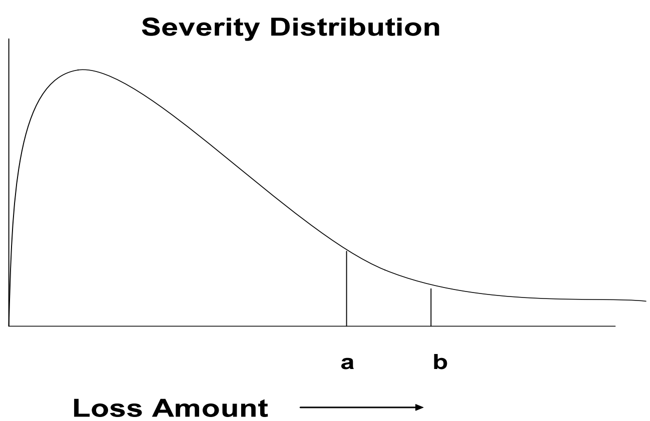
We devote the entirety of sections 1.1, 1.2 and 2 of our study to discussing the above interpretation. In a sense, it is a very simple insight but a key one in our modeling approach. In our study we have given an example to illustrate the method.

Justification of the Methodology

From sections 3.1 and 3.2 of our study:

In our methodology we model internal loss data by finding the best possible severity distribution and modeling the frequency as a Poisson distribution. Our methodology will also work if we choose to model frequency with other distributions.

The density function of the severity distribution model using internal (and when justified, external) loss event data will have the following shape:



The severity of the scenario data will be of the form [a,b] or others, as stated earlier. Looking at the preceding figure we can see that for the interval [a,b] we have a probability assigned by the density function. How do we interpret that probability? We say [a,b] is an event and the probability of that event is a forecast of the probability based on the internal and external loss experience (historical data) at an institution. So every event in the scenario has an associated "historical" forecast (as this forecast is simply based on historical data). The cornerstone of our approach is:

Important Assumptions

- Within a reasonably short period of time, all the losses that an institution will incur will come from the same family of distributions.
- The "hard work" of modeling historical data is to develop a model that we can trust for its ability to forecast within a reasonable period of time. Scenario data are used to adjust the tail structure of the model so that the forecast for the tail event can be improved.

We have discussed the justification for these assumptions, their usefulness, and their limitations in detail in section 4 of our study. In measuring operational risk, we are making a clear transition from a pure "data fitting" exercise to a more meaningful economic evaluation of the problem.

In order to improve the forecast from historical data, we adjust the distribution with another set of data, *i.e.*, scenario data. Scenario data also give the probability of a tail event. If the probability of that event in the scenario data is higher than the forecast of the probability of the same event given by the historical severity distribution, we account for the probability given in the scenario by adjusting the probability of the event to make it equal to what is given in the scenario data following the method given in section 3.1 and 3.2 in our study. If the probability in the scenario is lower than the corresponding forecast in the historical severity distribution, then we do not adjust the forecast of the historical severity distribution to match the probability given in the scenario. In our study this is what we meant when we said that we do not let the scenario data adjust the probability forecast of the historical distribution downward. An adjustment to lower the probability forecast implicit in the historical distribution would be contrary to the supplemental nature of scenario data, which are not to be considered a mitigant to the historical loss profile.

Suppose $f(\alpha, \beta)$ is the severity distribution based on the historical data. Let $S_1, S_2, \ldots, S_n$ be a set of scenarios with associated severities (we called them events) and forecasts of the probabilities of those severities (events). We are asking a simple question and our methodology is nothing but a method to answer the following question:

> *Given that the scenarios are the tail events, how much of the shape of the tail of $f(\alpha, \beta)$ needs to be adjusted to "match" the associated probabilities given in the scenarios?*

Essentially, we need to reweight the probability of various events using the scenarios. Sections 3.1 and 3.2 of our study provide a methodology to find those parameter values in a systematic way, given that the frequency distribution is modeled as Poisson with parameter λ. (Recall that the method given in sections 3.1 and 3.2 does not depend on which distribution is used to model the frequency distribution. It can be adjusted very easily if the frequency distribution is not Poisson.) The method in the working paper essentially reweights the probabilities of various (tail) events using the probabilities given in the set of scenario. This reweighting is not very trivial. The method in the working paper is carefully designed to take into account various issues that one may encounter while aggregating the cumulative effects of various scenarios in the set of scenarios.

In an earlier version of our study we used a uniform distribution for the supplemental data used in the method. I still believe that the choice of the uniform distribution was absolutely valid. However, based on valuable suggestions we received from several reviewers of our study we have changed our methodology with respect to the use of the uniform distribution. We now use the historical severity distribution for the supplemental data. As discussed in the section 4 of our study, very few data will be supplemented and those data will come from the historical severity distribution. Therefore suggestions that this method will result in a mixture distribution will be incorrect. Each data point used for the re-estimation of the parameter values will be IID.

Change of Measure

There have been one or two troubling comments that the claim in our study on the methodology as the Change of Measure (COM) method is incorrect. Here I demonstrate that our methodology is indeed a COM approach, the way this technique is generally understood in measure and integration theory. The texts we are citing here are very well known in this area, and they are very clear in supporting our claims.

Suppose the new estimated parameter values are α_1 *and* β_1. We call the resulting density function $f(\alpha_1, \beta_1)$ the implied density function (which is implied by the severities of the scenarios). In other words, using the method given in our study we reweight the probability of every event (a subset of R^+) from one probability measure $f(\alpha, \beta)$ to another $f(\alpha_1, \beta_1)$. This is consistent with the concept and representation of the Radon-Nikodym theorem given in the literature such as Arnold (1974). Etheridge (2002) clearly describes COM as a reweighting of the probability measure. One can obviously do this reweighting in several different ways. However every reweighting is driven by an objective. In our situation, we reweight the historical probabilities following the method given in sections 3.1 and 3.2 to make the probabilities of the tail events "match" with the probabilities given in the scenarios for those events. Many of us are familiar with the Girsanov theorem in the context of COM. Some of us often misunderstand that the Girsanov method is <u>the</u> change of measure method. Rather, we should note that the Girsanov method is nothing but a method for reweighting the probability of a nonstandard Brownian motion to make it a standard Brownian motion. Please refer to Etheridge (2002) for further explanation. Cerny (2009) discusses COM and its use in the context of asset pricing and gives a very thorough description for COM. Our work is consistent with the description and use of COM cited in Cerny (2009).

Once again, the method discussed in our study is essentially a reweighting of the probabilities of every event using the forecast of the scenarios. For some events the new probabilities may be the same as the old ones and for others they may even go down. Since the scenarios are primarily tail events, in the process of reweighting more probability will be attached to the tail and as a consequence less probability will be attached to the body of the density function so that the sum of total probabilities remains equal to one. On the other hand if the scenarios are primarily for the low severity events then in the reweighting process more probability will be attached to the body than in the tail. This is not a drawback of the method but rather reflects the flexibility of the method. The reweighting will happen in accordance with the data in the scenario which I would like to think will be for the tail events.

Also, to shift the mass from the tail of a distribution (in operational risk measurement the severity distributions are typically fat-tailed) to the body one would need a considerable amount of scenarios focusing on the body. That would be a contradiction to our assumption that most of the scenarios are for the tail events. A probability change at the tail is more important to the capital estimation in comparison to the changes in the body of the density function. Therefore under this approach we have always evidenced that the capital estimate goes up significantly.

We would like to think that the reweighting method used in our study is optimal under the economically meaningful assumptions made earlier and MLE method used to estimate the parameters. We are in the process of exploring the optimality condition further.

Since the method uses bootstrapping, our study gives a range for the parameter values. After experimenting with data from several financial institutions, we have found that standard errors (SE) of the parameter values in all of the cases are extremely low. This further empirically confirms the stability of the method.

Selection and Use of the Scenario Data

Scenario data are generated in a workshop. Out of many such data generated, an institution should decide how and which subsets of the scenario data are useful and important. As I said before, given the nature of the scenario data, its best use can be determined not only by a quantitative analysis but also with prudent qualitative judgment. In our study we also showed how one can evaluate the use of scenario data using COM numbers. We also showed how COM can help in the model selection process. In the following discussion I would like to point out some aspects of scenario data selection one should take into account.

From section 3 of our study:

The average probability of a scenario is $P = 1 \div (N \times \lambda)$. It is important to understand that it is an average probability only when we believe that the frequency is distributed as Poisson with parameter λ. The historical probability for an event [a,b] is given by $\mu = \int_a^b f(x)dx$.

If $P < \mu$ one may be tempted to ignore this scenario. One should note that the μ is an absolute probability in the sense that it is not dependent on the number of years or the choice for the frequency distribution function whereas the calculation of P is dependent on both of those elements. (Again, although the method in our study uses Poisson for frequency, it is not necessary that one adopt Poisson to use the method. The method can be suitably adjusted if one has any other assumption for the frequency distribution.) Also, when P is very close to μ but $P < \mu$, then it is not advisable to ignore the scenario without further quantitative or qualitative consideration. By all means $P < \mu$ is a "back-of-the-envelop" type of calculation to make an early prediction on the usefulness of the scenario but, once again, it should not be ignored blindly without further testing. In 2008 we implemented this selection method at a financial institution. We were surprised to note that the $P < \mu$ relationship cannot be taken for granted unless $P \ll \mu$ (P is significantly less as compared to μ). Therefore in the method suggested in our study we retest every scenario in the set of scenarios selected.

Following is another example where I show why one should be careful in using seemingly obvious reasoning for ignoring a scenario. Let us consider the following two scenarios within a unit of measure External Fraud (EF):

1. An event of [100million, 200million] can happen once in every 5 years: this scenario was generated for EF Retail Banking.

2. An event of [90million, 180million] can happen once in every 6 years: this scenario was generated for EF Custody Banking.

3. Frequency is distributed as a Poisson distribution with $\lambda = 20$

We should note that (1) and (2) are independent. Can we ignore (2) above and keep only (1) since event (1) severity at both ends is higher than the event (2) and the chance of happening (2) is lower than that of event (1) happening? The answer should be emphatically NO. Doing so is tantamount to saying that if EF at retail banking happens then the EF at custody banking cannot happen within the same period. Even though the chance of such an event happening at the same time is very low (once in 600 years on average using our notation), ignoring such an event will be inconsistent to our understanding of the tail events, *i.e.*, a tail event is a low probability and a high severity event. We should also note that in our assumption the events are independent but not mutually exclusive. In this case the severity of such a joint event is [190million, 380million], which is a significant severity.

In our method we do not make such "arbitrary" selections, which can have serious consequences in the capital calculations. When an institution is holding capital at a 99.9% level (or more) of the aggregate loss distribution, which when translated in terms of number of years is a once in a 1,000-year event, how can one ignore a loss that is a once in 600-year event? In our methodology we did consider this kind of selection process. We found it is not advisable to cherry pick losses on an arbitrary basis. The method of our study was designed with that view in mind and after experimenting with the data from several financial institutions.

When modeling internal data, one may have many choices for the severity distributions. In some cases the choice may not be very clear. In section 3.3 of our study we have discussed how scenario data could be used in the model selection process. Also, in section 3.3 we have discussed how one could decide on the usefulness of a scenario using the COM. One may need to consider the many other economic issues we discussed in our study before concluding on the choice of a scenario.

Other Issues

For the method in our study, we need the severity of the scenario event to be in a range such as [*a,b*] and others discussed earlier. We are aware of instances where scenario data are generated as a point estimate instead of a range (e.g. a $25 million loss occurring once in 25 years). Based on our empirical work we found that calculating a bound of ± 15%-20% around the point estimate will yield a severity range that is very comparable to data that are generated in a range format. In some sense, ± 15%-20% is an error bound for the point estimates. This suggestion was merely a stopgap measure until one can convert such point estimate into a range estimate. In section 4 of our study we discussed this issue in detail.

The method discussed in our study has been thoroughly tested for the last three and half years at several financial institutions with real scenario and historical data. All those assignments were done in the Insurance Economics and Risk Management group of Charles River Associates, International. I lead this group. We are a group of economists and risk management professionals with several years of experience in academia and in industry. We would not recommend to our clients any method that we had not validated ourselves. Validating this method is what we have been doing for the last three years, before we even attempted to write this working paper. We are now confident in our method and its theoretical underpinnings. Nonetheless, every model I have ever come across has its limitations and boundaries. We are also aware of the limitations of our method, which we discussed in detail in section 4 of our study.

Our work is fairly general in nature in the sense that it can handle data in any type of format as long as it comes in an interval. Even though we find neither the use of the format of the type [a, ∞] very useful nor we think that such data can be generated in a normal course out of a scenario workshop, our method can handle the data in that format also. We do not believe in creating data unnaturally in order to be compatible with a model. Our efforts have been to create a model that is compatible with the data and not other way.

Conclusion and a Response

I have attempted in this note to answer some of the key questions we received since March, 2010. The main objectives of this note are to:

- Explain the justification of the method we introduced,
- Explain why we think that the method is based on the Change of Measure approach, and
- Explain how the method suggested in our study uses and filters the scenarios properly.

I sincerely hope that our work will generate research interest in the area, as we consider scenarios to be an important piece in risk management. With that in mind <u>we did copyright</u> our work for the concepts and methodologies we introduced to protect our interests, but gave permission to everyone to freely use, <u>with proper citation</u>, the method and concept we introduced.

Recently I came across a working paper by Ergashev (2010). This work was put in the public domain in July of 2010. This note is not intended to critique the method and substance discussed in Ergashev (2010). However, I will discuss a few things that are directly related to our study which Ergashev (2010) misrepresented.

I found the claim in Ergashev (2010) that the method used there is "theoretically" based whereas ours was not, is <u>utterly misleading</u>. On further analysis, I have found that all the theories developed in section 3 of Ergashev (2010) essentially boil down to the following: the probability of a once in N-year event is $1 \div (N \times \lambda)$. <u>There is nothing else in his working paper that justifies such claim.</u>

To keep this note focused on our study, I am omitting here the derivation to $1 \div (N \times \lambda)$ following the chain of "theories" given in that working paper. The reader can easily follow very simple derivations. The first two sections (particularly section 1b) of Ergashev (2010) is a rehash of sections 1 through 3 of our study, yet I did not see any citation of our work there. This is unfortunate and against professional best practices.

I have come across several inaccuracies and inconsistencies in Ergashev (2010). The major ones (such as the algorithm for the selection of scenarios) I have addressed earlier in relation to the corresponding process given in our study. The comments made in section 5 of Ergashev (2010) in relation to our work are both misleading and inaccurate. I have attempted to clarify how that is so in this note.

We welcome new ideas and methods to extend our work. Such has been the history of research. Our work is certainly not the last word in solving the problem we attempted to solve, but we hope it is a useful step in the right direction.

References

Arnold, L, *Stochastic Differential Equations: Theory and Applications*. New York: John Wiley & Sons, 1974.

Cerny Ales, *Mathematical Techniques in Finance. 2nd ed.*, New York, NY: Princeton University Press, 2009.

Dutta, Kabir and David Babbel, "*Scenario Analysis in the Measurement of Operational Risk Capital: A Change of Measure Approach*" Working paper, Wharton Financial Institutions Center, The Wharton School, University of Pennsylvania 2010.

Ergashev, Bakhodir, "A Theoretical Framework for Incorporating Scenarios into the Operational Risk Modeling" Working Paper, The Federal Reserve Bank of Richmond, 2010.

Etheridge, A, *A Course in Financial Calculus*. Cambridge UK: Cambridge University Press, 2002

Kahneman, Daniel, Paul Slovic and Amos Tversky, *Judgment Under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press, 1982.