Scenario Analysis in the Measurement of Operational Risk Capital: A Change of Measure Approach[1]

Kabir K. Dutta[2]

David F. Babbel[3]

First Version: March 25, 2010; This Version: July 5, 2012

Abstract

At large financial institutions, operational risk is gaining the same importance as market and credit risk in the capital calculation. Although scenario analysis is an important tool for financial risk measurement, its use in the measurement of operational risk capital has been arbitrary and often inaccurate. We propose a method that combines scenario analysis with historical loss data. Using the Change of Measure approach, we evaluate the impact of each scenario on the total estimate of operational risk capital. The method can be used in stress-testing, what-if assessment for scenario analysis, and Loss Given Default estimates used in credit evaluations.

Key Words: Scenario Analysis, Operational Risk Capital, Stress Testing, Change of Measure, Loss Data Modeling, Basel Capital Accord.

JEL CODES: G10, G20, G21, D81

[1] We are grateful to David Hoaglin for painstakingly helping us by editing the paper and making many valuable suggestions for improving the statistical content. We also thank Ravi Reddy for providing several valuable insights and for help with the methodological implementation, Ken Swenson for providing guidance from practical and implementation points of view at an early stage of this work, Karl Chernak for many useful suggestions on an earlier draft, and Dave Schramm for valuable help and support at various stages.

We found the suggestions of Paul Embrechts, Marius Hofert, and Ilya Rosenfeld very useful in improving the style, content, and accuracy of the method. We also thank seminar participants at the Fields Institute, University of Toronto, American Bankers Association, Canadian Bankers Association, and anonymous referees for their valuable comments and their corrections of errors in earlier versions of paper. Any remaining errors are ours. Three referees from the *Journal of Risk and Insurance* provided thoughtful comments that led us to refine and extend our study, and we have incorporated their language into our presentation in several places.

[2] Kabir Dutta is a Senior Consultant at Charles River Associates in Boston. Kabir.Dutta.wg97@Wharton.UPenn.edu

[3] David F. Babbel is a Fellow of the Wharton Financial Institutions Center, Professor at the Wharton School of the University of Pennsylvania, and a Senior Advisor to Charles River Associates. Babbel@Wharton.UPenn.edu

Introduction

Scenario analysis is an important tool in decision making. It has been used for several decades in various disciplines, including management, engineering, defense, medicine, finance and economics. Mulvey and Erkan (2003) illustrate modeling of scenario data for risk management of a property/casualty insurance company. When properly and systematically used, scenario analysis can reveal many important aspects of a situation that would otherwise be missed. Given the current state of an entity, it tries to navigate situations and events that could impact important characteristics of the entity in the future. Thus, scenario analysis has two important elements:

1. Evaluation of future possibilities (future states) with respect to a certain characteristic.
2. Present knowledge (current states) of that characteristic for the entity.

Scenarios must pertain to a meaningful duration of time, for the passage of time will make the scenarios obsolete. Also, the current state of an entity and the environment in which it operates give rise to various possibilities in the future.

In management of market risk, scenarios also play an important role. Many scenarios on the future state of an asset are actively traded in the market, and could be used for risk management. Derivatives such as call (or put) options on asset prices are linked to its possible future price. Suppose, for example, that Cisco (CSCO) is trading today at $23 in the spot (NASDAQ) market. In the option market we find many different prices available as future possibilities. Each of these is a scenario for the future state of CSCO. The price for each option reflects the probability that the market attaches to CSCO attaining more (or less) than a particular price on (or before) a certain date in the future. As the market obtains more information, prices of derivatives change, and our knowledge of the future state expands. In the language of asset pricing, more information on the future state is revealed.

At one time, any risk for a financial institution that was not a market or credit risk was considered an operational risk. This definition of operational risk made data collection and measurement of operational risk intractable. To make it useful for measurement and management, Basel banking regulation narrowed the scope and definition of operational risk. Under this definition, operational risk is the risk of loss, whether direct or indirect, to which the Bank is exposed because of inadequate or failed internal processes or systems, human error, or external events. Operational risk includes legal and regulatory risk, business process and change risk, fiduciary or disclosure breaches, technology failure, financial crime, and environmental risk. It exists in some form in every business and function. Operational risk can cause not only financial loss, but also regulatory damage to the business' reputation, assets and shareholder value. One may argue that at the core of most of the financial risk one may be able to observe an operational risk. The Financial Crisis Inquiry Commission Report (2011) identifies many of the risks defined under operational risk as among the reasons for the recent financial meltdown. Therefore, it is an important financial risk to consider along with the market and credit risk. By measuring it properly an institution will be able to manage and mitigate the risk. Financial institutions safeguard against operational risk exposure by holding capital based on the measurement of operational risk.

Sometimes a financial institution may not experience operational losses that its peer institutions have experienced. At other times, an institution may have been lucky. In spite of a gap in its risk it didn't experience a loss. In addition, an institution may also be exposed to some inherent operational risks that can result in a significant loss. All such risk exposures can be better measured and managed through a comprehensive scenario analysis. Therefore, scenario analysis should play an important role in the measurement of operational risk. Banking regulatory requirements stress the need to use scenario analysis

in the determination of operational risk capital.[4] Early on, many financial institutions subjected to banking regulatory requirements adopted scenario analysis as a prime component of their operational risk capital calculations. They allocated substantial time and resources to that effort. However, they soon encountered many roadblocks. Notable among them was the inability to use scenario data as a direct input in the internal data-driven model for operational risk capital. Expressing scenarios in quantitative form and combining their information with internal loss data poses several challenges. Many attempts in that direction failed miserably, as the combined effect produced unrealistic capital numbers (e.g., 1,000 times the total value of the firm). Such outcomes were typical. As a result, bank regulators relaxed some of the requirements for direct use of scenario data. Instead, they suggested using external loss data to replace scenario data as a direct input to the model. External loss events are historical losses that have occurred in other institutions. Such losses are often very different from the loss experience of the institution. In our opinion, that process reduced the importance of scenarios in measuring operational risk capital. Previously, as well as in current practice, external loss data were and are used in generating scenarios.

We believe that the attempts to use scenario data directly in capital models have failed because of incorrect interpretation and implementation of such data. This work attempts to address and resolve such problems. Because scenarios have been used successfully in many other disciplines, we think that scenario data should be as important as any other data that an institution may consider for its risk assessments. Some may question, justifiably, the quality of scenario data and whether such data can be believable. We contend that every discipline faces such challenges. As we will show, the value in scenario data outweighs the inherent weaknesses it may have. Also, through systematic use we will be able to enhance the quality of the data.

In this paper we propose a method that combines scenario analysis with historical loss data. Using the Change of Measure approach, we evaluate the impact of each scenario on the total estimate of operational risk capital. Our proposed methodology overcomes the aforementioned obstacles and offers considerable flexibility. The major contribution of this work, in our opinion, is in the meaningful interpretation of scenario data, consistent with the loss experience of an institution, with regard to both the frequency and severity of the loss. Using this interpretation, we show how one can effectively use scenario data, together with historical data, to measure operational risk exposure and, using the Change of Measure concept, evaluate each scenario's effect on operational risk. We believe ours is the first systematic study of the problem of using scenario data in operational risk measurement.

In the next section we discuss why some of the earlier attempts at interpreting scenario data did not succeed and the weaknesses of current practices. We then discuss the nature and type of scenario data that we use in our models. Following that, we discuss our method of modeling scenario data and economic evaluation of a set of scenarios in operational risk measurement. We conclude with a discussion of some issues that may arise in implementing the method and of its use in other domains.

1. Problem Description

In their model for calculating operational risk capital, financial institutions subject to Basel banking regulations are required to use, directly or indirectly, four data elements:

- internal loss data (ILD), which are collected over a period of time and represent actual losses suffered by the institution;

[4] A basic source on these requirements is *Risk-Based Capital Standards: Advanced Capital Adequacy Framework — Basel II* at (http://edocket.access.gpo.gov/2007/pdf/07-5729.pdf)

- external loss data (ELD), which are loss events that have occurred at other institutions and are provided to the institution via a third-party vendor or from a data consortium;
- scenario data based on assessments of losses the institution may experience in the future; and
- business environment score, created from a qualitative assessment of the business environment and internal control factors (BEICF).

The regulatory rule does not prescribe how these elements should be used. However, given the similarity of operational losses to property/casualty losses, the measurement approach predominantly follows the loss distribution approach (LDA), which actuaries use for pricing property/casualty insurance.

Unit of measure is the level or degree of granularity at which an institution calculates its operational risk capital. The least granular unit of measure is enterprise-wide. More commonly, institutions calculate operational risk capital for several units of measure and aggregate those capital estimates. Units of measure are often determined by business line or type of loss event. Smaller business lines and/or less common types of loss events are frequently combined to create one unit of measure.

Of the four data elements, internal loss data are used primarily in the Loss Distribution Approach (LDA) to arrive at a base model. In that approach, one tries to fit two distributions: the severity distribution, which is derived from the amounts of all the losses experienced by the institution; and the frequency distribution, which is derived from the number of losses that have occurred at the institution over a predetermined time period (usually one year). As the frequency distribution, the Poisson distribution is the choice of nearly all financial institutions. Generally, the Poisson parameter is the average number of losses on an annual basis. A loss event (also known as the loss severity) is an incident for which an entity suffers damages that can be assigned a monetary value. The aggregate loss over a specified period of time is expressed as the sum $L_{Tot} = \sum_{i=1}^{N} L_i$, where N is a random observation from the frequency distribution, and each L_i is a random observation from the severity distribution. We assume that the individual losses L_i are independent and identically distributed, and each L_i is independent of N. The distribution of L_{Tot} is called the aggregate loss distribution. The risk exposure can be measured as a quantile of L_{Tot}. Dutta and Perry (2007) discuss the use and various challenges in modeling the severity distribution using internal loss data.

Given the characteristics and challenges of the data, an LDA approach resolves many issues. The sum L_{Tot} can be calculated by either Fourier or Laplace transforms as suggested in Klugman *et al*. (2004), by Monte Carlo simulation, or by an analytical approximation. We use the simulation method as well as an analytical approximation.

Regulatory or economic capital for the operational risk exposure of an institution is typically defined as the 99.9th or 99.97th percentile of the aggregate loss distribution. Alternatively, we can call it capital or price for the risk.[5]

1.1 Scenarios Are Not Internal Loss Data

Many financial institutions have been collecting internal loss data for several years. These data can be considered the current state for operational risk exposure. Additionally, many losses of various types and magnitudes have occurred only at other financial institutions. A financial institution may choose to evaluate such external loss data in order to understand the potential impact of such losses on its own risk profile. Typically, an institution analyzes those losses based on the appropriate magnitude and probability of occurrence, given the current state of its risk profile, and develops a set of scenarios.

[5] See footnote 4 for source.

Suppose institution A observes that institution B has incurred a $50 million loss due to external fraud, a type of operational loss. Institution A is also aware of the circumstances under which that loss occurred. After evaluating its own circumstances (current state), institution A determines that it is likely to experience a similar event once every ten years and that such an event would result in a $20 million loss. These are the frequency and severity of an event in the future state. Alternatively, the institution could specify a range, such as $15 million to $25 million, instead of a single number. We discuss this issue further in Section 2. Together with the description of the loss event, the specified severity and frequency constitute a scenario.

Suppose an institution has collected internal loss data for the last five years. It also generates a scenario for a certain operational loss event whose likelihood of occurring is once in ten years, resulting in a $20 million loss. It is inaccurate to interpret this as just another data point that could be added to the internal loss data. Doing so would change the scenario's frequency to once in five years from once in ten years. This **key insight** led us to develop a method that appropriately integrates scenario data with internal loss data. The problem most often reported from integrating scenario data with internal loss data is unrealistically large capital estimates. Because the integration process failed to consider the frequencies specified for the scenarios, the adverse scenarios were analyzed with inflated frequencies. A scenario could also be incorrectly analyzed with a frequency lower than specified. If the $20 million loss could occur once in 2.5 years, adding it to internal loss data from five years would dilute the effect of the scenario. A simplistic remedy would approximate the intended effect of this scenario by adding two such losses to the five years of internal data. Thus, frequency plays a key role in interpreting scenarios across financial institutions.

Suppose that, for the same unit of measure, two institutions have the same scenario of a $20 million loss occurring once in 10 years. One institution has an average of 20 losses per year, and the other has 50. For the institution with 20 losses per year, the scenario has much more impact than for the institution with 50 losses per year. Our method properly aligns the frequency of the scenario data with the time horizon of internal loss experience.

Continuing with the example of a $20 million loss whose frequency is once in ten years, in order to merge this scenario with internal loss data from five years' experience, we will have to consistently recreate internal data with a sample size equivalent to a period of ten years. Only then can we merge the scenario's $20 million loss with the internal data, and we would do so only if such a loss has not already been observed with sufficient frequency in those data. In other words, we use the current state of five years of observed internal loss data to generate enough data to determine whether the loss amount in the scenario is represented with sufficient frequency in the current severity distribution.

1.2 Measurement Practices Using Scenario Data

Rosengren (2006) adequately captured and summarized the problems with and the art of using scenario analysis for operational risk assessment. The issues discussed in Rosengren (2006) are still valid. In fact, since then, there has been very little, if any, focus on the development of scenario-based methodology for operational risk assessment. One exception was Lambrigger *et al.* (2007), who made an early attempt to combine expert judgment with internal and external operational loss data. Their informal approach was to make qualitative adjustments in the loss distribution using expert opinion, but they provided no formal model for incorporating scenarios with internal and external loss data. The methods that we found in the literature are very *ad hoc*, and most integrate scenarios and internal or external data without sound justifications.

One method[6] pools the severities from all the scenarios and then samples from that pool a severity for each unit of measure that the institution is using for internal or external loss data modeling. In each replication of the simulation, the severities are sampled according to the probabilities assigned to the scenarios for that unit of measure. If a severity is chosen, it is added to other severities chosen in that replication. If no severity is chosen, zero is added. From the observed distribution of the summed severity amounts (over the trials), the 99.9th or 99.97th percentile is chosen. This number is then compared with the corresponding percentile of the loss distribution obtained using internal or external loss data, and the institution must decide which number to use for regulatory capital. Typically the scenario-based number will be much higher than the number based on internal or external data. In such situations, a number between the two is chosen as the 99.9th or 99.97th percentile. Rarely, the scenario-based 99.9% or 99.97% level number would be added to the corresponding number obtained using internal or external loss data to provide an estimate of extreme loss. This method suffers from the drawback that the universe of potential severe loss amounts is limited to the severity values assigned to the scenarios, which are completely isolated from internal and external loss data. This approach closely resembles sampling from an empirical distribution. Dutta and Perry (2007) highlight some of the problems involved.

Another method derives two types of severity numbers from the one scenario created per unit of measure. One figure is the most likely severity outcome for the scenario, and the other represents the worst severity outcome. Then a purely qualitative judgment is made to interpret these two severity values. The worst-case severity outcome is put at the 99th percentile (or higher) of the severity distribution obtained from internal or external data for that unit of measure, and the most likely severity outcome is put at the 50th percentile of the severity distribution. The 99.9th or 99.97th percentile is obtained from the loss distribution after recalibrating the severity distribution with these two numbers. As in the previous method, the resulting percentile is compared with the corresponding percentile of the distribution based on internal or external loss data. Typically the institution uses purely qualitative judgment to choose an operational risk capital amount between the two figures.

All other methods of which we are aware are variations or combinations of these two. Institutions adopt some type of *ad hoc*, often arbitrary, weighting of the 99.9th or 99.97th percentiles from the loss distributions derived from both internal loss event data (sometimes also including external loss event data) and the scenario data to arrive at a final model-based regulatory or economic capital number.

2. Generating Scenario Data

External loss data are the primary basis for scenario generation at every financial institution. Several sources offer external data.[7] Those data contain the magnitude of the loss amount and a description of the loss, including the name of the institution, the business line where the loss happened, and the loss type. Basel regulatory requirements categorize operational losses into seven types: Internal Fraud, External Fraud, Employment Practices and Work Place Safety, Client Products and Business Practices, Damage to Physical Assets, Business Disruptions and System Failures, and Execution Delivery and Process Management.

Prior to generation of scenarios, risk management decisions determine the unit of measure, which often crosses loss types or loss types within business lines. It could also cross business lines or sub-business lines. For some units of measure, internal loss experience may not be adequate to support any meaningful analysis. Some financial institutions supplement such units of measure with external data. From

[6] The methods described are not published but observed in practice. Financial institutions have implemented similar methods.

[7] The First database from Fitch is one good source of data. It is based upon publicly available information on operational losses with severities exceeding $1 million that have occurred at financial institutions.

preliminary research we have undertaken on external data, we are not comfortable using our approach on units of measure that have insufficient internal loss data to develop a meaningful and stable model. Although our method does not explicitly depend on which data are used for calibration, an unstable base model will give poor estimates of the effects of scenarios. Thus, we often form an "other" category that includes adequate internal loss data.

To generate scenarios within a unit of measure, an institution uses a scenario workshop, typically conducted by a corporate risk manager or an independent facilitator. The participants are business line managers, business risk managers, and people with significant knowledge and understanding of their business and the environments in which it operates. Workshop participants discuss the business environments and current business practices, and take guidance and help from external data such as the following:

Event A

At bank XYZ, when selling convertibles to clients, an employee makes inappropriate promises to buy them back at a certain price. Market conditions move in the wrong direction, and the bank is required to honor the commitment. As result, the bank suffers a loss of $200,000,000.[8]

Question for Workshop: Could a similar situation occur at our institution? If so, what is the potential magnitude of the loss, and how frequently might this happen?

The unit of measure for this event will usually be Client Products and Business Practices (CPBP). After considering a range of possibilities, participants agree on a scenario related to this event. We are assuming one scenario per incident type. Multiple scenarios should be carefully combined without sacrificing their value.

The unit of measure such as CPBP can be thought of as a process driven by many factors, such as unauthorized employee practices in selling convertibles. A scenario is not loss data. It is an impact and sensitivity study of the current risk management environment. The data in a scenario have two quantitative components – severity and frequency – and one descriptive component – the type of loss within the unit of measure. The description identifies the type of scenario within a process and is an essential characteristic of a scenario. In the above example, the scenario is for unauthorized employee practices of selling convertibles within CPBP. Often more scenarios are generated in a workshop than will be useful for quantification. In that situation scenarios may be filtered, taking into account their descriptive components. This decision is best made at the scenario generation workshop. In the above example the risk management team should very carefully decide whether a scenario on unauthorized employee practices for selling equity can be ignored when a scenario of unauthorized employee practices for selling convertibles was also generated, even though both are "unauthorized employee practices" within the larger event class of CPBP.

The severity in a scenario can be a point estimate (e.g., $2 million) or a range estimate (e.g., between $1 million and $3 million). We prefer to work with range estimates, intervals of the form $[a, b]$, as we believe that in such a hypothetical situation a range captures the uncertainty of a potential loss amount. This choice is consistent with the continuous distributions we use for modeling the severity of internal loss data. A continuous distribution assigns positive probability to ranges of values, but zero probability to any individual value. We can convert a point estimate to a range estimate by setting the lower and upper bounds of the range at appropriate percentages of the point estimate. We revisit this choice in Section 4.

[8] This example was supplied to us by a banking associate. Our understanding is that it is an adaptation of an event from an external database.

The frequency in a scenario takes the form $m \div t$, where m is the number of times the event is expected to occur in t years. We interpret m as the number of events that we expect to occur in a sample of size $n_1 + \ldots + n_t$, where n_i is the number of losses observed annually, sampled from the frequency distribution of the internal loss data for that particular unit of measure. We assume that the capital calculation is on an annual basis. Stating the frequency denominator as a number of years allows us to express the sample size as a multiple of the annual count of internal losses at an institution. Like the severity, the frequency could take the form of a range such as $m \div [t_1, t_2]$. For a range we interpret $m \div t_1$ as the worst-case estimate and $m \div t_2$ as the best-case estimate. Alternatively, one could take a number between t_1 and t_2, such as their average. We are making a subtle assumption that we use throughout the analysis.

Assumption 1: During a short and reasonably specified period of time, such as one year or less, the frequency and severity distributions based on the internal loss data for a unit of measure do not change.

This assumption is important because our methodology is conditional on the given severity and frequency distributions (in this case, based on internal loss data). Justification for the one-year time threshold lies in the loss data collection process and capital holding period at major financial institutions in the USA. To interpret the assumption in terms of time and state of riskiness of the unit of measure, we would say that at time zero (today) we have full knowledge of the loss events for the unit of measure. Using this knowledge, we forecast the future for a reasonable period of time in which we can safely assume that the assumption is valid. We stress that scenario data are not the institution's loss experience. Our analysis does not use scenario data as a substitute for internal loss data. Scenario data represent the possibility of a loss; we are proposing a method to study its impact. Therefore, we make another vital assumption.

Assumption 2: The number of scenarios generated for a unit of measure is not more than the number of internal loss events observed in that unit of measure.

Subjectivity and possible biases will always be inherent characteristics of scenario data. Methods for interpreting scenarios must take these features into account. As Kahneman, Slovic, and Tversky (1982) put it: "A scenario is especially satisfying when the path that leads from the initial to terminal state is not immediately apparent, so that the introduction of intermediate stages actually raises the subjective probability of the target event." We have undertaken research that seeks to explain how one could control and reduce the biases and subjectivity in scenario data in the context of operational risk. Very preliminary results show that scenario data generated in the format discussed above are less subjective and therefore more suitable than data produced in other formats. We also think that this is the most natural format in which workshop participants can interact and express their beliefs. The discussion of how judgment happens in Chapter 8 of Kahneman (2011) further corroborates some of our findings in the context of scenario data generation for operational risk.

As noted earlier, in current practice, scenario data are predominantly influenced by external data. We have several concerns. External data are historical events that have occurred at other institutions. Generating scenario data by solely using or over-relying on external data and without considering many other hypothetical situations may defeat the purpose of the scenario. We will discuss these issues in later studies, as detailed discussions of them are beyond the scope of this paper.

3. Methodology

For the method to work effectively, it is important that the current state (severity and frequency distributions using historical losses[9]) be estimated accurately. We may have many candidates for the severity distribution, and often the choice is not clear. In such situations it will be advisable to use all models that could possibly be a good fit, particularly when the institution did not experience many losses for a particular unit of measure. This approach will ensure more stability in the measurement of the current state, which is an important cornerstone for this method. The frequency distribution, on the other hand, provides the information to determine how many losses would need to occur before one would expect to observe a loss of a given magnitude. In order to take account of sampling variability, one may need many thousands of replications, drawing an N from the frequency distribution and then drawing N losses from the severity distribution.

From the current state, we would like to predict the probability of an event's happening in the future. On the other hand, scenario data also implicitly predict the probability of an event, which we define as a subset of the positive real numbers, usually an interval. More than likely, these two probabilities will not match. Therefore, the probability distribution of the current state must be adjusted in such a way that it accounts for the scenario probability. This step is necessary in order to calculate the equivalent distribution if the scenario loss actually happened with the frequency specified in the scenario, proportional over the specified period of time. If the scenario data lower the probability compared with the historical loss data, then for practical reasons we do not alter the probability predicted by the historical loss data.

Borrowing terminology from the Black-Scholes option pricing concept, we call the probability distribution implied by the scenario the *implied probability distribution*. The probability distribution for the current state, estimated from historical losses, is the *historical probability distribution*.

The scenario events' historical probabilities could be based on internal and external loss data. In our experiments, however, probabilities based on internal loss data have proved to be much more stable than those based on both internal and external loss data. Our method is based on two primary assumptions:

Assumption 3: Within a reasonably short period of time, all the losses that an institution incurs come from the same family of distributions.

Assumption 4: The challenge in 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 improve the model's forecasts for tail events.

We discuss the justification for these assumptions, their usefulness, and their limitations in detail in Section 4. 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.

Suppose $f(x|\theta)$[10] is the density function of the severity distribution based on historical data. $f(x|\theta)$ could be any distribution, including a mixture or any other suitable combination. Discussion of the choice and appropriateness of a distribution for fitting the internal loss data is beyond the scope of this work.[11] Let

[9] From now onward we will denote by historical loss data the internal and when appropriate the supplemental external loss event data chosen by the institution for a particular unit of measure.

[10] The generic parameter θ may have as many components (individual parameters) as the distribution requires.

[11] Dahen et al. (2010), Dutta and Perry (2007), Hoaglin (2010), and Nagafuji (2011) are good sources for such a discussion.

$S_1, S_2, \ldots, S_r$ be a set of scenarios, independently occurring,[12] with associated events and implied probabilities. Our methodology aims to answer the following question:

> *Given that the scenarios are tail events, how much does $f(x|\theta)$ need to be adjusted so that its probabilities for those events "match" the probabilities implied by the scenarios?*

Essentially, we revise the probabilities of various events to take the scenarios into account. Sections 3.1 and 3.2 show how to adjust the parameter values in $f(x|\theta)$.

The method does not depend on the form of the frequency distribution. The conventional choice is the Poisson distribution, but the method could easily use other distributions. The revision of the probabilities of multiple events, independently occurring, is not always simple. The method is designed to take into account various issues that may arise from the combined effects of multiple scenarios.

In practice, each event is a bounded interval (i.e., its endpoints, a and b, are finite). And, because the severity distributions are continuous, P_{ab}, the probability assigned to the event, does not depend on whether a and b are in the interval. Thus, we define the range of the event as $b - a$. We say an event $[a, b]$ has *occurred* when a sample contains an observation x such that $a \leq x \leq b$. When $a \to b$, $P_{ab} \to 0$. In other words, when the range of an event goes to zero, its probability of occurring (frequency) also goes to zero. We use this fact when discussing sensitivity analysis due to the range of an event.

Suppose the new estimated parameter (vector) is θ_1. We refer to the resulting density function, $f(x|\theta_1)$, as the *implied density function* (implied by the scenarios). In other words, we reweight the probability of every event. Etheridge (2002) describes Change of Measure as a reweighting. In that sense the method is essentially a Change of Measure method. One can accomplish the reweighting in various ways, but every reweighting is driven by an objective. Our method reweights the historical probabilities to make the probabilities of tail events "match" the probabilities implied by the scenarios.

Because the scenarios are primarily tail events, the process of reweighting will move probability from the body of the severity distribution to the tail. If the scenarios consisted primarily of low-severity events, the reweighting process would move probability from the tail to the body. The method has the flexibility to handle such sets of scenarios. In order to move substantial weight from the tail to the body, however, one would need a considerable number of low-severity, high-frequency scenarios. In that hypothetical situation the set of scenarios would not be considered economically realistic.

The reweighting method should be optimal under the economically meaningful assumptions made earlier and the maximum-likelihood method of estimating the parameters. We also evaluate the effect of each scenario on the Change of Measure in order to understand its economic impact.

Omitting the dependence on parameters, we let $f(x)$ be the historical probability density function for the severity for the particular unit of measure.

If we randomly draw n observations from $f(x)$, the expected number of observations between a and b is $n \times P_{ab}$. For a fixed value of n, the actual number of observations will reflect sampling variability. Overall variability in the number of observations will also involve the variability that arises in sampling n from the frequency distribution. For the random variable N representing the number of losses in a year, we denote the probability function by $\lambda(n)$, the historical frequency distribution. In other words, the probability that N takes the value n is $\lambda(n)$.

[12] We assume the independent occurrence of events given by the scenarios. This is consistent with loss distribution approach. Also, we have seen very little evidence of the dependence of operational risk events in the internal and external loss data.

We refer to probabilities based solely on historical data interchangeably as a measure of current state, current probabilities, or historical probabilities.

In theory, the scenario-based $f(x)$ and $\lambda(n)$ can have a different shape than the ones obtained using ILD. However, as we discuss in our application, the scenario-data-based $f(x)$ and $\lambda(n)$ will differ from their ILD-based counterparts only in the values of their parameters.

3.1 Calculation of Implied Probability Distributions

An example in the Appendix illustrates the method discussed here. We first consider a single scenario, whose event has a severity given as $[a, b]$ and a frequency in the form of $m \div t$, which means that m such events are likely or expected to occur in a period of t years. In adjusting the probability distribution to take the scenario into account, we revise the historical probability distribution so that the number of occurrences of the event in a sample equivalent to *t* years of losses is equal to m. We obtain the size of such a sample by drawing *t* observations, n_i (i = 1, …, *t*) from the historical frequency distribution, yielding a total of $n_{Tot} = \sum_{i=1}^{t} n_i$ losses.

We then draw a sample of n_{Tot} observations from the historical severity distribution. Suppose that this sample contains k occurrences of the event $[a, b]$. If k is less than m, then we draw $m - k$ additional observations from the historical severity distribution restricted to the interval $[a, b]$ and combine them with the initial n_{Tot} observations**.** We then fit a severity distribution to the combined data set by adjusting the parameters of the severity distribution. The resulting severity distribution is the *implied severity distribution* due to the scenario in which $[a, b]$ occurs m times in t years.

In practice, n_{Tot} will be substantially larger than $m - k$. Therefore it should be satisfactory to work within the same family of distributions and re-estimate the parameters of the historical severity distribution using the combined data set. We revisit this issue in Section 4.

In addition to the scenario, *S*, the implied severity distribution depends on the value of n_{Tot} and on the particular sample from the historical severity distribution. For simplicity, however, we denote the implied severity distribution by $I(x|S)$. More generally, if $\psi = \{S_1, S_2, \dots, S_r\}$ is a set of scenarios, then $I(x|\psi)$ denotes the implied severity distribution due to the combination of scenarios in ψ. The key to appropriately deriving the combined effect is to preserve the frequencies of the events in those scenarios.

Calculation of $I(x|\psi)$ is very similar to the case of one scenario. Scenario S_i in ψ has a severity range $[a_i, b_i]$ and frequency $m_i \div t_i$. Thus, each scenario may reflect a different time span. We take T to be $\max\{t_i\}$. If event S_i occurs m_i times in t_i years, then in T years we should have $m_i(T/t_i)$ occurrences of S_i. We are making a simple linear extrapolation. One could use a stochastic extrapolation based on the frequency distribution. For a Poisson frequency distribution, we find that linear extrapolation gives a very close approximation to stochastic extrapolation. We have normalized the frequencies to a common time span. The result for S_i may not be a whole number, but we define the normalized frequency $c_i = m_i(T/t_i)$ and use it in the calculations that take into account overlaps among events. The calculation below is based on the independent occurrence of the scenarios in the set of scenarios. In the Appendix, we give a probabilistic justification of the method.

In the case of more than one scenario, it is necessary to take into account the frequency of each scenario. In a simple example, if two scenarios have exactly the same severity range, then (on average) in a sample of n_{Tot} losses the number of occurrences of losses in that severity range must be equal to or greater than the sum of the frequencies indicated by the two scenarios. More commonly, the severity ranges overlap,

but do not coincide, and more than two scenarios may be involved. In such situations, we adjust the c_i in a way that reflects the overlaps. Thus, for each scenario we have two cases:

1. The severity range of the scenario is disjoint from the ranges all other scenarios.
2. The severity range of the scenario overlaps, completely or partially, with the ranges of other scenarios in the set ψ.

Case 1

When the severity range of a scenario is disjoint from the ranges of all other scenarios, its normalized frequency remains unchanged.

Case 2

When the severity range of a scenario overlaps fully or partially with the severity ranges of other scenarios, we arrange the scenarios in order and calculate a cumulative frequency for each scenario.

We sort the set of scenarios so that the lower bounds of their severity ranges are in non-decreasing order: $a_1 \leq \cdots \leq a_r$. In our application, the value of the lower bound is never less than zero.

To calculate a cumulative frequency for each scenario in the set ψ, we construct a *lower triangular* r by r matrix, R, whose (i, j) element is r_{ij} ($i = 1, \ldots, r, j = 1, \ldots, r$) with $r_{ij} = 0$ when $j > i$.

For $i = 1, \ldots, r$, we define $r_{ii} = 1$.
For $i = 2, \ldots, r$ and $j = 1, \ldots, i–1$,

$r_{ji} = (b_i - a_j) \div (b_i - a_i)$ when the j^{th} scenario overlaps with the i^{th} scenario and $b_j \geq b_i$
$= (b_j - a_j) \div (b_i - a_i)$ when the j^{th} scenario overlaps with the i^{th} scenario and $b_j < b_i$
$= 0$ otherwise.

We denote the column vector of the unadjusted normalized frequencies by $C = (c_1, \ldots, c_r)'$. Then the (adjusted) cumulative frequencies are given by the column vector: $\tilde{F} = RC = ([f_1], \ldots, [f_r])'$, where $[f]$ denotes the result of rounding f to the nearer integer. These cumulative frequencies allow the algorithm to follow the order of the lower bounds, taking each scenario separately, and determine whether the sample from the severity distribution (perhaps augmented by additional observations for preceding scenarios) contains the required number of occurrences of the scenario's event.

For $i = 1, \ldots, r$, the sample of n_{Tot} losses should contain f_i occurrences of the event associated with scenario S_i. If it contains fewer than f_i occurrences, we augment the sample by drawing the appropriate number of additional losses from the interval $[a_i, b_i]$.

The preceding steps are based on a single sample. To obtain a stable estimate of the effect of the set of scenarios, we repeat the process at least 10,000 times. For each such realization of n_{Tot}, the initial sample of losses, and the additional losses, we determine the implied severity distribution $I(x|\psi)$. Then for each $I(x|\psi)$ we denote by $\eta\{I(x|\psi)\}$ the 99.9% or 99.97% quantile of the loss distribution obtained by using $I(x|\psi)$ as the severity distribution and the historical $\lambda(n)$ as the frequency distribution.

We could also use $I(x|\psi)$ to calculate other measures. We are using the historical frequency distribution, but the frequency also changes. However, since the amount of data augmentation is small relative to the total size of the sample, the change in the average value of the frequency will be negligible. If we needed to adjust the frequency distribution, the adjusted Poisson parameter would be equal to (n_{Tot} + total number of additional draws from the historical severity distribution for all scenarios combined)$\div T$. To

save time, the 99.9% or 99.97% level can be calculated using the single loss approximation formula given by Böcker and Klüppelberg (2005) instead of a Monte Carlo simulation of one million trials. We take the median of all the $\eta\{I(x|\psi)\}$ to arrive at the final capital number due to scenarios in set ψ conditional on the current state and the corresponding $I(x|\psi)$ is the implied probability distribution that was used for capital calculation, in this case the median of all 10,000 $\eta\{I(x|\psi)\}$.

Here we are trying to find the median implied distribution due to the combined effect of a set of scenarios under certain criteria, such as the 99.9% level of the aggregate loss distribution resulting from the implied distribution. Alternatively, one can form a 95% band of the estimates by ignoring the bottom and top 2.5% of the estimates, or one can take the median of all the estimates as a final number. If we take the 95% band, then there will be two corresponding distributions for this band. If we take an average, there may not be an exactly corresponding $I(x|\psi)$. In that case, we take the nearest one to represent the implied distribution.

3.2 Economic Evaluation of Scenarios: The Change of Measure

Each scenario will change the historical probability for a given severity range. The Change of Measure (COM) associated with the scenario is given by:

$$\textit{Change of Measure} = \frac{\textit{Implied probability of the severity range}}{\textit{Historical probability of the severity range}}$$

This value is a one-step conditional COM, and it applies only to the severity distribution. In general, one can make a similar calculation for the frequency distribution. In our application, however, the change in the frequency distribution will have negligible impact on overall computations and scenario studies. The step is from Time 0 (or current state of loss, i.e., ILD) to Time 1, the period for which we are trying to estimate the operational loss exposure. This time period should not be long (usually no more than a year). If the time period is long, one may have to re-evaluate the current state. This same issue arises in the financial economics of asset pricing. It is extremely difficult, if not impossible, to price an option of very long maturity given the current state of the economy. Borrowing the language of financial economics, we are trying to determine the states by the possible future values of the losses. For tail events we do not know what that loss will be. Therefore it is impossible to know the future values of tail event losses. We are trying to estimate a range of values for likely outcomes. We need to consider the possibilities for changes regarding events that will cause the tail losses.

The unit-free nature of the COM makes it useful for evaluating the loss distribution derived from historical data. Suppose we have two distributions that adequately fit the historical loss data. For one, the COM is 20, and for the other it is 5. We can say that the second distribution is a better predictor of the given scenario event than the first, without knowing anything about those distributions.

The COM quantifies the impact of a particular scenario and its marginal contribution in risk valuation. An extremely high COM can be due to any of these three situations:

1. The historical measure is inaccurate and inconsistent with the risk profile of the institution;
2. The scenario is nearly impossible given the current state of the institution;
3. The risk of the institution is uninsurable (self-retention or through risk transfer).

In the third situation above, the risk may be uninsurable by an institution itself, but it may be insurable collectively among all institutions with similar risk exposure. However, that may not be possible without causing a systemic risk. The pros and cons of such a situation are very well discussed in Cummins,

Doherty, and Lo (2002). An operational risk scenario should also be developed and considered in light of this important aspect discussed in that paper.

On the other hand, a low COM may indicate redundancy of a scenario. Before we interpret its numerical value, however, we must ensure that the numerator and denominator of the COM are calculated correctly. If a probability distribution involving scenarios is not based on the historical distribution, the resulting values of the COM will be incorrect or misleading. And if the historical distribution is not evaluated correctly, both the implied distribution and the COM will be incorrect. The method discussed in Section 3.1 systematically updates the historical distribution to produce the implied distribution.

The COM can also be used to evaluate our valuation of the current state. If we had several possible distributions to fit to the historical loss severity data, the COM could be a criterion for model selection using the scenarios that we believe are a good representation of the future state of the loss profile.

Therefore, using scenario analysis and the COM, one can validate both the current state and the usefulness of the scenario under consideration. The method described here can be used to evaluate one scenario at a time or any subset of scenarios. In asset pricing one often uses the price of an option to predict the future distribution. For example, in the seminal work of Jackwerth and Rubinstien (1996) probability distributions of the future price of underlying assets were estimated from the prices of sets of options. Using the same reasoning and the COM, we could reassess our choice of a severity distribution in terms of its predictive power.

3.3 Example

The results shown and discussed in this section are based on actual internal loss event data and scenario data made available by a major financial institution. Operational loss data as well as scenarios for operational losses are highly sensitive proprietary information. We do not have permission to provide descriptive statistics for the data in our examples.

Table 1 shows the results of goodness-of-fit tests for five distributions used to model the internal loss data for a particular unit of measure at a financial institution. No external loss event data were used in modeling the severity and frequency distributions.

Table 1: Results of Two Goodness-of-Fit Tests for Five Distributions

Column headings are defined as follows: Stat: test statistics, CV: critical value (at 95% confidence level), P: p-value, H = 0: null hypothesis, i.e., a distribution fits the data, could not be rejected; H = 1: alternative hypothesis, i.e., the distribution does not fit the data, could not be rejected.

Distributions	Anderson-Darling				Kolmogorov-Smirnov			
	Stat	CV	P	H	Stat	CV	P	H
Lognormal	0.701652	0.744959	0.061618	0	0.059954	0.066526	0.111536	0
GPD	0.434731	0.755586	0.310396	0	0.038805	0.063516	0.637776	0
Loglogistic	0.390558	0.672978	0.318861	0	0.048933	0.058910	0.208124	0
Burr	0.268320	0.509786	0.442925	0	0.040082	0.055923	0.423949	0
GB2	0.272043	0.485646	0.311828	0	0.041333	0.055685	0.269032	0

We show two goodness-of-fit tests: Anderson-Darling (AD) and Kolmogorov-Smirnov (KS). Any of the five distributions could be used to model the internal loss data. However, on the basis of the *p-value* of the AD test, Burr has the best fit, and GB2 (the parent distribution of Burr), GPD, and loglogistic will be

among the next best choices. Lognormal gives the worst fit. On the basis of the *p-value* of the KS test, GPD is best, and Burr is second best. If we want to put more emphasis on the tail than on the body, we should give priority to the AD test. Then Burr will be the distribution of choice for this unit of measure. All five distributions produced a reasonable 99.9% and 99.97% level capital with the internal loss data.

Table 2 depicts the sixteen scenarios. For this unit of measure, the institution has experienced many severe losses (tail events) and therefore generated many tail scenarios to understand the institution's risk exposure. To disguise the actual data, we have coded the lower bound of the first scenario as 1 and expressed the upper bound of the first scenario and the lower and upper bounds of the other scenarios as a multiple of the lower bound of the first scenario. Table 2 also shows the COM of each scenario under the five distributions in Table 1. Table 3 shows the COM when the set of scenarios includes only one scenario at a time. Using the methodology of Section 3.2 for computing the implied distribution, we use cumulative frequency for Table 2 and normalized frequency for Table 3. Use of normalized frequency for computing COM in Table 3 enables us to compare the COM among all the scenarios on a uniform basis.

Table 2 shows the value of COM for each scenario as a combined effect (or *group effect*) of all the scenarios, whereas Table 3 shows the value of COM as a *pure effect* or an *individual effect* of a scenario. The difference in COM between the group effect and individual effect will be pure group effect for each scenario. Because of the construction of our method the individual effect cannot be less than one. Numbers less than one in Table 3 are due to simulation variation.

Table 2 (Group Effect): A Set of Scenarios for a Particular Unit of Measure: Event Range, Cumulative Frequency, and Change of Measure under Five Distributions

The table depicts the sixteen scenarios used. To disguise the actual data, we have coded the lower bound of the first scenario as 1 and expressed the upper bound of the first scenario and the lower and upper bounds of the other scenarios as a multiple of the lower bound of the first scenario. The set of scenarios for calculating the implied probability distribution includes all sixteen scenarios. Therefore, the table shows the value of COM for each scenario as a combined effect (or group effect) of all the scenarios.

No	Lower Bound	Upper Bound	Cumulative Frequency	Change of Measure				
				GPD	Loglogistic	Lognormal	Burr	GB2
1	1	5	7	0.88	0.96	0.92	0.88	0.95
2	2	18	14	1.01	1.13	1.13	1.05	0.97
3	13	26	18	1.41	1.39	1.56	1.53	1.37
4	18	33	16	1.55	1.44	1.72	1.68	1.55
5	60	110	10	2.34	1.69	2.81	2.50	2.79
6	75	200	20	2.66	1.76	3.29	2.80	3.35
7	75	225	34	2.70	1.77	3.35	2.83	3.43
8	75	250	54	2.73	1.78	3.39	2.87	3.49
9	76	336	60	2.82	1.80	3.51	2.95	3.66
10	103	206	60	2.86	1.81	3.66	2.99	3.73
11	106	152	63	2.75	1.79	3.49	2.89	3.53
12	119	186	58	2.90	1.82	3.76	3.03	3.82
13	400	600	3	4.42	2.08	7.19	4.37	7.11
14	1160	1935	20	6.54	2.35	14.49	6.11	12.77
15	1697	1979	8	7.07	2.40	17.08	6.53	14.34
16	3500	7500	7	10.02	2.67	33.97	8.76	24.17

From the upper and lower bound columns of Table 2 and Table 3 we can make an estimate of the relative severities of the events in the scenarios. For example, the last three scenarios are 1160, 1697, and 3500 times as severe as the first scenario. The severity of first scenario is in the upper 10% of the internal loss experience of the institution. Therefore, these scenarios are much more severe in comparison with the institution's internal loss experience.

Table 3 (Individual Effect): A Set of Scenarios for Particular Unit of Measure: Event Range, Normalized Frequency, and Change of Measure under Five Distributions

The table depicts the sixteen scenarios used. To disguise the actual data, we have coded the lower bound of the first scenario as 1 and expressed the upper bound of the first scenario and the lower and upper bounds of the other scenarios as a multiple of the lower bound of the first scenario. The set of scenarios for calculating the implied probability distribution includes only that individual scenario. Therefore, the table shows the value of COM for each scenario as a pure effect (or individual effect).

No	Lower Bound	Upper Bound	Normalized Frequency	Change of Measure				
				GPD	Loglogistic	Lognormal	Burr	GB2
1	1	5	7	1.00	1.00	1.00	0.98	1.02
2	2	18	10	1.00	0.97	0.99	1.06	1.02
3	13	26	14	0.98	0.95	1.00	1.05	0.98
4	18	33	7	0.99	0.99	1.00	0.97	0.98
5	60	110	10	1.02	1.02	0.99	0.98	0.90
6	75	200	13	1.01	0.99	1.04	0.99	1.05
7	75	225	14	0.98	1.02	1.02	0.97	1.02
8	75	250	20	1.07	1.04	1.13	1.11	1.26
9	76	336	7	1.01	0.99	1.00	1.02	1.10
10	103	206	14	1.14	1.01	1.14	1.16	1.22
11	106	152	5	1.00	0.97	1.02	1.06	0.96
12	119	186	3	0.98	1.04	1.03	0.98	0.97
13	400	600	3	1.06	1.03	1.16	1.04	1.26
14	1160	1935	20	2.97	1.49	4.63	2.67	5.76
15	1697	1979	2	1.16	1.06	1.23	1.11	1.49
16	3500	7500	7	2.03	1.16	2.82	1.71	4.11

Selection of a scenario is an important aspect in the use of scenarios. More often than not, the workshop generates many scenarios without understanding their usefulness in measuring risk. For scenarios with very high severity values (extreme tail events) the group effect may be more severe than the individual effect, whereas for scenarios with comparatively low severity values, the individual effect may be dominant. For the first twelve scenarios, the values of the COM for all distributions show very similar individual effects and very similar group effects. Pending more-detailed analysis, we believe for these scenarios the individual effect dominates. For the last four scenarios, the group effect clearly dominates.

In order to understand the effect of a particular scenario (as in a *what-if* analysis), we could evaluate that scenario alone, in the absence of the other scenarios. In Table 3 we can see that none of the individual effects of any scenario is particularly high for any distribution. Modeling of current state using Burr or loglogistic could reasonably predict the future states given in any of the scenarios individually, especially the extreme tail events. More than likely, not all of the scenarios will happen during a specified period. In that case it is more important to understand the individual effects of tail events. On the other hand, the

group effect can be used for *stress testing* the current states using collections of scenarios. Here Tables 2 and 3 together can help in understanding the use and purpose, and selection of a set of scenarios. The first twelve scenarios with comparatively low severity inflated the COM, at least by a multiple of two, for the last four scenarios.

Alternatively, we could remove each scenario in turn and compare the COM for the remaining subset against the COM for the full set. Information on the group effect and individual effect for each scenario will be of value in understanding the risk sensitivity of various factors that constitute the risk for a unit of measure. Chart 1 is a graphical representation of Table 3 (Panel 1) and Table 2 (Panel 2).

The five distributions generally give different results for those scenarios. The loglogistic shows the least variation. In that sense, loglogistic has the best predictive power for the given set of scenarios, conditional on the internal loss data. On the other hand, lognormal and GB2 performed worst in this respect. The GB2 distribution has a fatter tail than the loglogistic distribution. Also the GB2 has four parameters, whereas the loglogistic has two. Thus, a distribution's predictive power for an extreme event is not necessarily determined by either its tail shape or its number of parameters. Therefore, for this set of scenarios and the internal loss experience of the institution for this unit of measure, loglogistic is the best choice for modeling the severity of the data.

One could extend the analysis further in the context of the three situations discussed in Section 3.2. Concluding that loglogistic is the best choice (and not the Burr) with respect to the scenarios is tantamount to believing that Situation 1 applies – i.e., that the historical measure is inaccurate and inconsistent with the risk profile of the institution. On the other hand, if we believe that the Burr distribution accurately models the risk profile of this unit of measure, then the last three scenario events will probably be less likely to occur in the set of all sixteen scenarios, given the current state and risk profile of the institution with respect that unit of measure; that is, we would believe that Situation 2 applies.

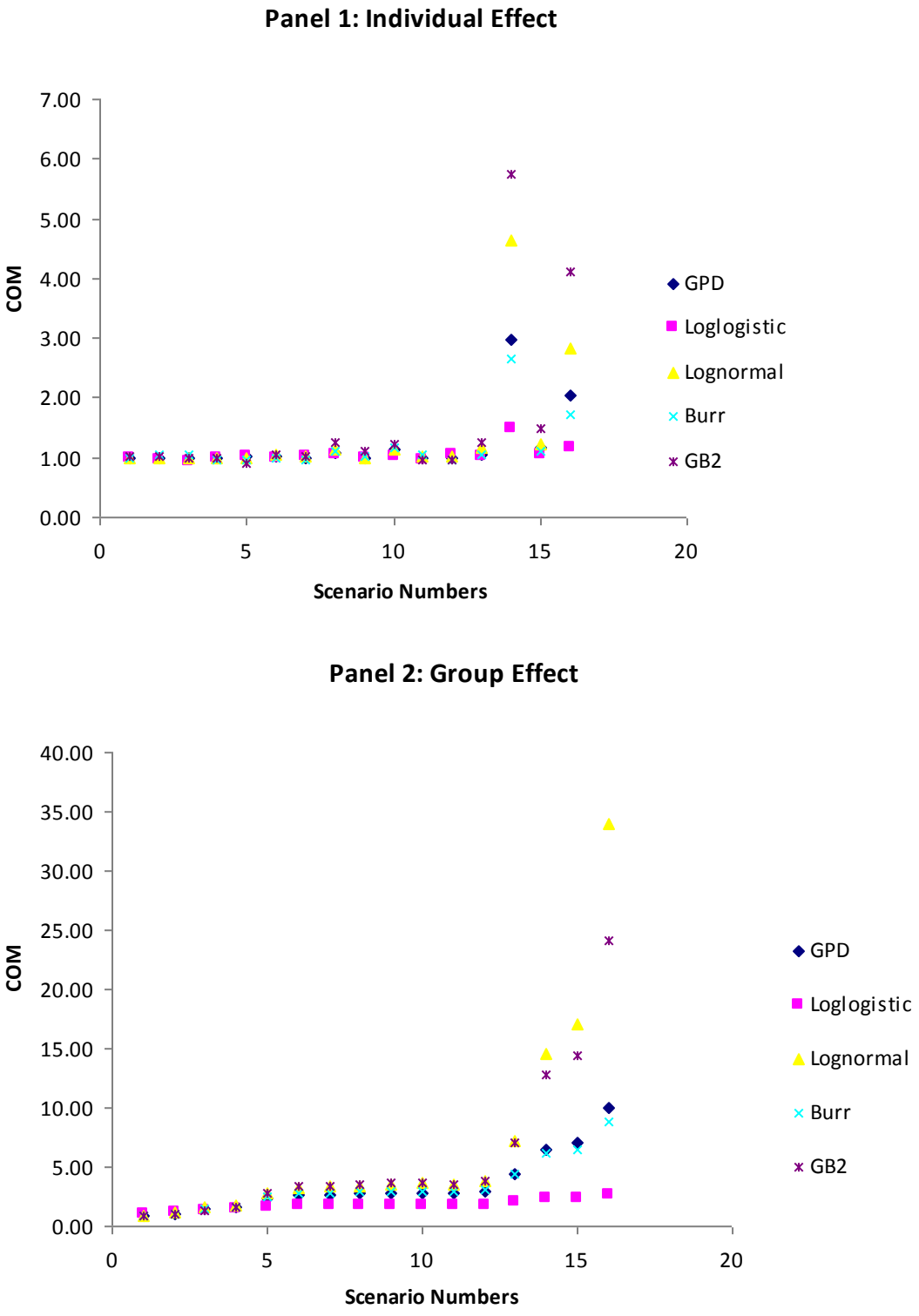
The third possibility is that we trust the severity model for the internal loss data using the Burr distribution and also believe that the last three scenario events are equally possible in the set of all sixteen scenarios. If we then arrive at an unrealistic capital estimate, we would consider the last three scenarios uninsurable (Situation 3). In that situation, we would revisit the risk controls and address the risk management problem through risk control mechanisms instead of through additional capital holdings. The present application of COM can easily be adapted for insurance pricing of an individual or group risk.

Table 4: Relative Standard Error of the Parameter Estimates

Distribution	Param-1	Param-2	Param-3	Param-4
Lognomal	0.2859%	1.2741%		
GPD	1.7569%	0.0002%		
Loglogistic	0.2868%	2.3786%		
Burr	8.9288%	1.8110%	5.2130%	
GB2	10.1283%	8.3559%	13.7386%	10.4954%

Table 4 shows the relative standard error (also known as the coefficient of variation) of the parameter estimates from 10,000 trials in the simulation used to merge the scenario data with the internal loss data across various units of measure. Here we are showing the standard error for the best fitting distribution for a particular unit of measure. The relative standard error is obtained for each parameter of a distribution by calculating standard deviation of the parameter values and dividing it by the mean of the parameter value. The parameter estimates are very stable under this approach. We have tested this with the data from many different financial institutions and found this always to be the case. For the GB2 distribution we observed the greatest sensitivity to the parameter values.

Chart 1: Change of Measure for the 16 Scenarios When the Five Distributions Are Used to Model the Internal Loss Data



4. Discussion

Historical loss data are at the core of the methods discussed here. If the historical-loss-data-based model for either severity or frequency is not estimated correctly, then these methods will suffer from instability and inaccuracy. For evaluation of severity and frequency models, we strongly suggest that, along with goodness-of-fit tests for each model, one should judge performance by the following criteria, used in Dutta and Perry (2007):[13]

1. Realistic - If a method fits well in a statistical sense, does it generate a loss distribution with a realistic capital estimate?
2. Well-Specified - Are the characteristics of the fitted data similar to the loss data and logically consistent?
3. Flexible - How well is the method able to reasonably accommodate a wide variety of empirical loss data?
4. Simple - Is the method easy to apply in practice, and is it easy to generate random numbers for the purposes of loss simulation?

In addition, one should stress-test each distribution in terms of coherence, consistency, and robustness. A log-log plot may be a useful tool to ascertain the threshold of a Pareto tail if the Pareto family of distributions is used to fit the tail of the distribution. It could reduce the number of simulations. Institutions have often used scenarios either in the absence of internal loss data or when very little internal loss data had been observed. Our method cannot be used with scenario data alone.

The method is flexible enough that an institution can use more than one scenario for each unit of measure, but that flexibility has limits. Heuristically, if the number of scenarios for a unit of measure is greater than 5-10% of its annual number of historical loss events, the implied probability distribution may start to show instability. Thus, one should not use very many scenario data points to augment the historical-loss-data-based severity distribution. Our experience has shown that if the scenario set has only short-term events (between 20 and 25 years), then we rarely augment the historical sample by more than 4% of its sample size. On the other hand, if the scenario set has several long-term events (more than 50 years), then we need to augment the sample by less than 1% of its sample size. For example, for a long-term (100 year) event in a unit of measure that has an average annual loss count of 250, we will have a sample size on the order of 25,000. As D'Agostino and Stephens (1986) note, for such large data sets, goodness-of-fit tests for any distribution are expected to fail unless by stroke of luck every data point happens to come from that specific distribution. Also, as we insisted, a robust fit of historical data will ensure that the fit is still statistically non-rejectable with the augmented sample. In Appendix C we show the Q-Q plots of the tail area (95^{th} percentile and above) of the scenario-augmented data we used to make sure the effect of the scenario data was properly accounted in the capital calculation. If any of the above conditions are not met, then the data have to be retested for implied probability. In fact, in those situations it will be better not to use this method than go through data fitting of an impractical amount of data such as 25,000 observations.

In theory, our method does not prove that the form of distribution that fits the historical severity data is still a good fit for the augmented data set. Ideally, one would make a fresh assessment of the fit of the distribution with the augmented data. Integration of scenario losses may make the tail of the distribution significantly fatter than the one based on the historical data. In that sense it may make the refitting of the distribution more challenging. One may potentially encounter a situation in which no known distribution or parsimonious mixture is a good fit. In those situations, as we suggested earlier, one may want to

[13] Page 58 of the *Observed Range of Practice in the Key Elements of Advanced Measurement Approaches (AMA)* issued in July 2009 and page 39 of *Operational Risk- Supervisory Guidelines for the Advanced Measurement Approaches issued in June 2011* by the Basel Committee on Banking Supervision recognizes these criteria as good practice for model selection.

evaluate the validity and appropriateness of the scenario. From a practical point of view, however, refitting the distribution often has only a negligible effect on its shape. Also, we are making an assumption that the historical-loss-data-based severity and frequency distributions are good predictors of the future, as long as the forecast horizon is not too long. Scenario data usually pertain to tail events that have not been observed to a reasonable extent in the severity distribution (based on historical data alone) to take appropriate tail shape. The number of scenario data points is expected to be much smaller than the number of historical data points. The nature and purpose of such data is to forecast the frequency of an event within a specified period. The severity distribution also forecasts the frequency of such an event within the same period. In the COM approach we make sure that the frequency of an event forecast from the historical-data-driven severity distribution is as great as the frequency given by the scenario, if we believe such a scenario may be a possibility. Therefore, we recalculate the distribution parameters after augmenting the sample with the appropriate amount of scenario data.

Another reason for not changing the family of the historical distribution is the evolving nature of scenario data generation. Our approach places more value on historical loss data, especially since many institutions have been collecting historical loss data for a considerable period of time.

Because prioritizing scenario data over historical data has many practical negative implications, our method is conditional on historical loss data. The loss distribution derived from historical loss data serves as the base case. Accordingly, for an event given in a scenario, our approach never adjusts the historical probabilities downward to the probability based on the scenario frequency. The only adjustments are upward. If, as is typical, scenarios are used to predict tail events rather than body events, the process shifts probability toward the right, making the tail of the distribution fatter.

In our method, the range, $[a,b]$, of a scenario event has a finite upper bound, but the method can be modified when b is infinite. For people generating scenario data, ranges will be a convenient way to express the severity. If severity data come in another form and we can map those data into an interval, we can use our method. In some instances scenario data are generated as a point estimate instead of a range (e.g., a $25 million loss occurring once in 25 years). In empirical work not discussed here, we found that bounds of ±15%-20% around the point estimate will yield a severity range that is very comparable to data generated in a range format. Instead of using ±15%-20% across all scenarios, it is better to use a tighter bound (say 15%) for near-term events and progressively increase the percentage for longer-term events. This procedure is consistent with the subjective nature of a scenario. In some sense we are accounting for the subjectivity as the number of years increases for the occurrence of the scenario.

One should be careful in arbitrarily enlarging or reducing the interval range. In Section 3 we observed that, when the range goes to zero, the corresponding probability (equivalently, the frequency) under a continuous severity distribution also goes to zero. Conversely, if the range is too large, the probability becomes high. If we keep the frequency constant and arbitrarily narrow the range, we will get a high but incorrect estimate. In such situations the frequency should also be reduced to maintain consistency with the narrower range. On the other hand, if we increase the range arbitrarily without correspondingly increasing the frequency, we will get a low but incorrect estimate. This feature is very important for conducting a sensitivity study of this method and also for interpreting the scenario data.

Our method is simulation-based and can be used for studying the impact of one scenario at a time or the combined effects of a set of scenarios. Therefore, it can be used for stress-testing the historical-data-based model, for what-if analysis under a set of scenarios, and for measuring the total impact of all scenarios in conjunction with historical loss data for estimating operational risk. The method can used in removing redundancies in a set of scenarios. Peura and Jokivuolle (2004) use a simulation-based approach to stress-test the capital adequacy of a financial institution with respect to credit risk using scenarios for business

cycles, among others. The business cycle scenario can also be studied in the context of Loss Given Default (LGD) using the approach proposed here.

COM also has important uses for communication. Using COM, we can express the net effect of a scenario in relation to the current loss experience at an institution and for a particular unit of measure. The COM evaluation is important in itself. Evaluating scenarios based on COM provides a clear understanding of the likelihood of the scenario in relation to all other events and conditional on the current loss experience. This probability study of the event in a scenario can help to create a financial product for risk transfer contingent on the occurrence of that event. A financial product traded in the market is essentially a promise of a future cash flow with an objective probability. The price (at Time 0) of the financial product traded in the market is calculated by converting the objective probability to a risk-neutral probability and then discounting it with a risk-free rate. One could do the same for the creation of a catastrophe bond (cat bond) or for other insurance products using the probability obtained for the events underlying each scenario. A further study in this domain will be useful for possible risk transfer of some of the operational risks classified by each scenario.

Our method assumes independence among scenarios. However, if the scenarios in a set are correlated, our method can easily be extended to account for correlation among scenarios within a unit of measure. Since our actual scenario data were not generated in a correlated way, we were unable to use our method to illustrate our work for correlated scenarios. We leave it for future work when such data will be available.

5. Conclusion

The study of market risk of financial products uses the Change of Measure as an important evaluation criterion. Using COM, one can properly evaluate the possibility of a future cash flow of an instrument. This principle was the motivation behind our work. We asked how a scenario could be evaluated using the Change of Measure approach. In a sense, future cash flows are a scenario for some economic process. For us, it is the loss generation process driving the severity distribution calibrated by historical loss experience. If operational risk scenario data had quality similar to market risk scenario data (such as option, future, and forward contracts with high liquidity), we could use scenario data to calibrate our loss generating process, as in many market risk modeling approaches, such as in Jackwerth and Rubinstein (1996). For operational risk, we propose to use COM in validating and updating the loss generating process. As problematic as operational risk scenario data are, the use of scenario data for operational risk measurement should have the same priority as internal and external loss data. MacMinn (2005) differentiated corporate risk exposure between pure and speculative. In that framework operational risk would be classified as a pure risk. Similarly, we can differentiate among various tail risks and their impacts on the value of the firm based on their insurability. An interesting extension of this work would be to analyze various tail risks using COM in the MacMinn framework. Another interesting extension would be to employ the method presented here with scenarios, developed through a thoughtful economic analytical approach, whose frequency and severity data have been vetted and adjusted through a Bayesian or other suitable approach.

Through construction of implied distributions, we have shown how to merge historical and scenario data in a systematic way. We have thoroughly tested our method using actual data from several financial institutions. From those implementations and tests, we see the use and further development of this method through research for many applications beyond the measurement of operational loss exposure of an institution. Through this method we found a good use for scenario data that were thought to be practically impossible to model in a systematic way. We would not characterize scenario data as forward-looking. We believe that in the context of operational risk measurement, all data elements that the regulatory world has identified are equally important. One needs to find a systematic and meaningful way to extract the

information contained in those data elements. Our method is an effort in that direction. The practice of operational risk measurement has some unfortunate tendencies to blame the data if they cannot be modeled with available knowledge, as if the data must be good for the model to work and not the reverse.

If we can draw any inference from the experiences reported in other disciplines that use scenarios as a cornerstone for measuring and managing uncertainty, we can say that in due course the data will become slightly (but not a whole lot) better. We should search for a method to compensate for the inherent problem of data quality. We believe our approach, evaluating scenarios using a Change of Measure approach, is a modest step in that direction.

So far, research on estimating operational risk has heavily emphasized statistical searches for the "best" severity distribution. In our opinion, it should also be viewed as a problem in financial economics. If in the end institutions must hold billions of dollars as operational risk capital, its measurement should involve tools and techniques from financial economics. We hope this work will stimulate research interest in studying operational risk in the context of economics, rather than as a pure (and very often "impure") data fitting exercise. Also, we hope that this work will encourage effective use of scenarios in measuring "pure" risk in the framework of MacMinn (2005) such as operational risk and other types of risk covered by an insurance contract. Although we have illustrated our method using operational risk data, our approach is more general. We think it will be applicable in the management decision making process, where scenarios are used to adjust, stress-test, or assess the predictive power of a statistical distribution of an underlying process.

References

Basel Committee on Banking Supervision, "Observed Range of Practice in Key Elements of Advanced Measurement Approaches (AMA)." July 2009.

Basel Committee on Banking Supervision, "Operational Risk- Supervisory Guidelines for the Advanced Measurement Approaches" June 2011.

Böcker, Klaus and Claudia Klüppelberg, "Operational VaR: a closed-form approximation." *Risk*, December, 2005.

Cummins, J. David, N. Doherty, and A. Lo, "Can insurers pay for the 'big one'? Measuring the capacity of the insurance market to respond to catastrophic losses." *Journal of Banking and Finance* 26(2-3): March 2002.

Dahen, Hela, Georges Dionne and Daniel Zajdenweber, A practical application of extreme value theory to operational risk in banks, *The Journal of Operational Risk* 5(2): Summer 2010.

D'Agostino, Ralph B. and Michael A. Stephens, *Goodness-of-Fit Techniques*. New York, NY: Marcel Dekker, Inc., 1986.

Dutta, Kabir and Jason Perry, "A Tale of Tails: An Empirical Analysis of Loss Distribution Models for Estimating Operational Risk Capital." Working Paper, Federal Reserve Bank of Boston, 2007.

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

Financial Crisis Inquiry Commission, "Final Report of the National Commission on the Causes of the Financial and Economic Crisis in the United States." United States Public Affairs, January 2011.

Hoaglin, David C., "Extreme-Value Distributions as g-and-h Distributions: An Empirical View." *2010 Proceedings of the Annual Meeting of the American Statistical Association*. Alexandria, VA: American Statistical Association, 2010.

Jackwerth, J.C. and M. Rubinstein, "Recovering Probability Distributions from Option Prices." *Journal of Finance* 51(5): December 1996.

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

Kahneman, Daniel, *Thinking Fast and Slow*. New York: Farrar Strauss and Giroux, 2011.

Klugman, Stuart A., Harry H. Panjer, and Gordon E. Willmot, *Loss Models: From Data to Decisions*, second edition. New York: John Wiley & Sons, Inc, 2004.

Lambrigger, D., P. Shevchenko and M. Wüthrich, "The Quantification of Operational Risk using Internal Data, Relevant External Data and Expert Opinions." *Journal of Operational Risk*, 2(3): 2007.

MacMinn, Richard D., "On Corporate Risk Management and Insurance." *Asia-Pacific Journal of Risk and Insurance*, 1(1): June 2005.

Mulvey, John, M. and Hafize G. Erkan, "Risk Management of a P/C Insurance Company Scenario Generation, Simulation and Optimization." 2003 Proceedings of Winter Simulation Conference.

Nagafuji, Tsuyoshi, "A Simple Formula for Operational Risk Capital: A Proposal Based on the Similarity of Loss Severity Distributions Observed among 18 Japanese Banks." Working Paper, Bank of Japan, 2011.

Office of the Comptroller of the Currency, Federal Reserve System, Federal Deposit Insurance Corporation, and Office of Thrift Supervision, "Risk-Based Capital Standards: Advanced Capital Adequacy Framework — Basel II." *Federal Register* 72(235): 2007 pp. 69, 288-69, 445.

Peura, Samu and E. Jokivuolle, "Simulation based stress tests of banks' regulatory capital adequacy." *Journal of Banking and Finance* 28(8): August 2004.

Rosengren, Eric, "Operational Risk Scenario Analysis Workshop: Scenario Analysis and the AMA." 2006. Presentation at the Bank of Japan, Tokyo.

Appendix

A. An Illustration of the Methodology

The following example illustrates the steps outlined in Section 3. The example given here is based on an actual implementation of the Change of Measure method at a financial institution. To protect the proprietary information of the institution, the real data have been altered. Therefore, the example here should be used purely for understanding the steps outlined in Sections 3.1 and 3.2 and not for any other interpretation of either data or results.

Table A1 presents a sample of scenario analysis data. The frequency of the scenario should be interpreted as once in the stated number of years. The financial institution obtained range estimates for the severity of each event in scenario workshops; therefore severity is represented by upper and lower severity bounds. Had the workshops yielded point estimates, we would have calculated 15–20% intervals around them.

Table A1: Scenario Data

No	Frequency	Lower Bound	Upper Bound
1	25	$ 29,412	$ 147,059
2	20	$ 73,529	$ 294,118
3	5	$ 88,235	$ 588,235
4	10	$ 500,000	$ 1,200,000
5	62	$ 1,168,500	$ 1,291,500

Normalization of the Scenario Frequencies

Table A2 shows the normalized frequencies. Each normalized frequency represents the number of times a scenario would occur in the number of years equal to the maximum frequency in the scenario data (in this example, 62 years).

Table A2: Scenario Data with Normalized Frequencies

No	Normalized Frequency	Lower Bound	Upper Bound
1	2.48	$ 29,412	$ 147,059
2	3.1	$ 73,529	$ 294,118
3	12.4	$ 88,235	$ 588,235
4	6.2	$ 500,000	$ 1,200,000
5	1	$ 1,168,500	$ 1,291,500

Calculation of Cumulative Frequency for Overlap Scenarios

If the severity ranges of the scenarios overlap, the cumulative frequencies account for the overlap. We want to determine whether the model specified by historical data can satisfy all scenarios within a given severity range, not just each scenario within the range individually. The rationale for our approach is based upon the assumption that the scenarios are independent.

In Table A1 the scenarios are already in order of increasing lower bounds. The percentage by which Scenario 2 overlaps Scenario 1 is (147,059 – 73,529)/(147,059 – 29,412) = 63%. Table A3 shows (as percentages) the overlap matrix (denoted by *R* in Section 3.2) for the five scenarios.

The cumulative frequency for Scenario 2 is the sum of the normalized frequency for Scenario 2 and the product of the percentage overlap between Scenario 2 and Scenario 1 and the normalized frequency for Scenario 1:

$$\text{Cumulative Frequency for Scenario 2} = 3.1 + (63\%)(2.48) = 4.65\,, \tag{1}$$

which becomes 5 when rounded.

Because Scenario 3 overlaps Scenarios 1 and 2, its cumulative frequency includes contributions from those two scenarios.

Since we have sorted the scenarios so that the lower bounds of their severity ranges are non-decreasing, the cumulative frequency for a scenario needs to include contributions only from overlapping scenarios that precede it in the ordering. The normalized frequencies in these calculations are those implied by the scenarios; they are not necessarily the same as those implied by the historical loss data. The process of comparing the cumulative frequencies of the scenarios against the numbers of occurrences in the sample of n_{Tot} losses from the historical severity distribution takes the scenarios in the sorted order. Thus, any additional losses sampled for preceding scenarios will already be included in the augmented sample.

The rightmost column of Table A3 shows the cumulative frequencies for the five scenarios.

Table A3: Scenario Data Overlap Matrix and Cumulative Frequencies

No	Lower Bound	Upper Bound	Normalized Frequency	1	2	3	4	5	Cumulative Frequency
1	$ 29,412	$ 147,059	2.48	100%	0%	0%	0%	0%	2
2	$ 73,529	$ 294,118	3.1	63%	100%	0%	0%	0%	5
3	$ 88,235	$ 588,235	12.4	50%	93%	100%	0%	0%	17
4	$ 500,000	$1,200,000	6.2	0%	0%	18%	100%	0%	8
5	$1,168,500	$1,291,500	1	0%	0%	0%	5%	100%	1

Calculation of Implied Distribution

In deriving the implied distribution from the historical distribution, we want to add only those events that do not occur often enough in the sample data. If the cumulative frequency specified by the scenario data is the same or less than the number of events in the sample loss data over the same severity range, then the scenario is satisfied by the sample data. If the cumulative frequency specified by scenario data is greater than the number given by sample loss data, then the scenario is not satisfied by sample data and additional events must be added to the sample data. We calculate the difference between the observed frequency from the sample data, k, and the expected scenario cumulative frequency, f. If k is less than f, we add $f - k$ events, drawn from the severity range for the scenario in the historical severity distribution.

For example, in Table A3 the cumulative frequency for Scenario 4 is eight events in the range $500,000 to $1,200,500. If the sample loss data (augmented for Scenarios 1, 2, and 3) included only three events in that range, then five events would be drawn in the range of $500,000 to $1,200,500 from the historical severity distribution and added to the sample. Conversely, if the sample loss data had eight or more events

over that range, then Scenario 4 would already be satisfied by the historical loss data, and no additional loss events would be added to or subtracted from the sample data.

We re-estimate the parameters of the loss severity distribution that we used to generate the sample data. The augmented sample represents both historical and scenario data. We would expect that most losses in this vector are generated historical data points, and only a few are augmented to satisfy the cumulative frequencies of the scenarios. As discussed in Section 4, if this condition does not hold, then our method may not be suitable for the set of scenarios.

We repeat this process 10,000 times to reduce the simulation variance of the sample data generation process and estimate a range of parameters of the loss severity distribution. We also recalculate here, although it may not be necessary, the frequency parameter, accounting for the "uncovered" events that are added to the sample data. At each step of the simulation process we could also calculate the 99.9% or 99.97% level value of the aggregate loss distribution using the re-estimated parameters of the frequency and severity distributions. One can use Monte-Carlo simulation of one million trials for that or, in order to save time, one could use the Single Loss Approximation formula given in Böcker and Klüppelberg (2005). Out of all the losses reaching the 99.9% or 99.97% threshold, in each trial generated we use the median for the purpose of estimating operational risk given the historical loss experience and the scenario generated at an institution.

Table A4 shows the severity ranges for the five scenarios, together with the cumulative frequencies from Table A3. The fifth column of Table A4 shows the historical frequency count, which is the number of loss occurrences within the severity range in a sample from the severity distribution used to model the severity of the historical loss event data.

Table A4: Cumulative Frequency vs. Historical Frequency

No	Lower Bound	Upper Bound	Cumulative Frequency	Historical Frequency
1	$ 29,412	$ 147,059	2	320
2	$ 73,529	$ 294,118	5	104
3	$ 88,235	$ 588,235	17	5
4	$ 500,000	$ 1,200,000	8	2
5	$ 1,168,500	$ 1,291,500	1	1

B. Probabilistic Explanation of the Method[14]

Let us assume that severities of the losses are distributed according to distribution F and frequencies are Poisson distributed with mean frequency λ.

Given an event with probability of occurrence p, what is the expected number of samples one has to draw in order to observe it? If n draws are needed, then there would be n–1 failures before observing a success. This implies that the number of draws has a geometric distribution with mean $1/p$. On the other hand, because this loss occurs on average once every M years, the expected number of losses necessary to observe it is λM. Immediately, it follows that

[14] Ilya Rozenfeld helped in developing this section.

$$p = \frac{1}{\lambda M} \tag{1}$$

Now, Suppose a scenario is given where it is expected that the loss exceeding amount *L* would occur once every *M* years. Let the event be a loss greater than *L*, it follows that $p = P(X>L) = 1–F(L)$. Therefore, $1/(1–F(L))$ can be interpreted as the expected number of losses that need to occur in order to observe one loss greater than *L*.

Then using (1), $p = 1 - F(L) = \frac{1}{\lambda M}$

We assume that the scenarios are independent. Let $E_1, E_{2,\dots,} E_n$ be a set of events from the corresponding scenarios $S_1, S_{2,\dots,} S_n$ independently occurring with probabilities $P_1, P_2,\dots, P_n$ then the probability of one or more of the events happening is given by:

$Prob\ (E_1 \cup E_2 \cup \dots \cup E_n) = \sum_1^n Prob(E_i) - \sum_{i<j} Prob(E_i \cap E_j) + \sum_{i<j<k} Prob(E_i \cap E_j \cap E_k) + \dots$

$+ (-1)^{n+1} Prob\ (E_1 \cap E_2 \cap \dots \cap E_n)$ (2)

By independence of events and from step (1)

$Prob\ (E_i \cap E_j) = Prob(E_i)\ Prob\ (E_j) = \frac{1}{\lambda^2 M_i M_j}$ where E_i and E_j happen on an average every $M_i\ and\ M_j$ years, respectively.

Similarly

$Prob\ (E_i \cap E_j \cap E_k) = Prob(E_i)\ Prob\ (E_j)\ Prob\ (E_k) = \frac{1}{\lambda^3 M_i M_j M_k}$

$Prob\ (E_i \cap E_j \cap \dots \cap E_n) = Prob(E_i)\ Prob\ (E_j) \dots Prob\ (E_n) = \frac{1}{\lambda^n M_1 M_2 \dots M_n}$

If $\lambda, M_1, M_2, \dots M_n \geq 10$ then $\frac{1}{\lambda^2 M_i M_j} \leq 1/10{,}000$, $\frac{1}{\lambda^3 M_i M_j M_k} \leq 1/1{,}000{,}000\dots$

and $\frac{1}{\lambda^n M_1 M_2 \dots M_n} \leq 1/1{,}000{,}000$ for $n \geq 3$

In our application $\lambda, M_1, M_2, \dots,$ and M_n are typically greater than 10 . On a rare occasion we observed that *M* may be less than 10 but in those cases λ is very high. Therefore we choose to ignore all the terms except the first term of equation (2) without compromising the accuracy of the resulting probability. Therefore, $Prob\ (E_1 \cup E_2 \cup \dots \cup E_n) = \sum_1^n Prob(E_i)$ for all practical purposes of our application. The method developed in Section 3 ensures the sum of probability of events given in the set of scenarios. By doing so we ensure that our probability calculation is accurate up to the fourth decimal place.

C. Q-Q Plots for the Tail Plots of the Scenario-Augmented Data

In the discussion section we noted that the scenario-augmented data may be too large for any meaningful goodness-of-fit test. However, to make sure that scenario data (which usually adjust the tail area of a distribution) were fitted well, we checked this in two ways. First, we used the concept of the *Change of Measure* to evaluate the validity of the scenarios. Second, we also used Q-Q plots to statistically evaluate the goodness-of-fit of the scenarios and ensure that their effects were adequately captured in the capital calculation. In the following chart we show the Q-Q plot corresponding to the median of the 10,000 trials we used for the capital calculation of the scenario-augmented data. Here we can see that the Burr distribution, which was our best choice for the fitting the internal loss data, is still a good fit in the tail area for the scenario-augmented data. However, the loglogistic distribution, which was our second best choice for fitting the internal loss data, happens to fit the tail of the scenario-augmented data better than the Burr. Here one can choose the loglogistic distribution for internal loss and the scenario data after the economic evaluation of the scenarios made using the *Change of Measure* approach discussed in Section 3.

Q-Q Plots for Burr and Loglogistic Distributions for the Scenario-Augmented Data (at the level of 95th percentile and above)