THE Open GROUP

Calculating Reserves for Cyber Risk

Using Calibrated Estimates for VaR and CVaR Calculations with Open FAIR™ Risk Analysis

A White Paper by:

Mike Jerbic and Dr. Robert Mark

May 2024

Calculating Reserves for Cyber Risk:
Using Calibrated Estimates for VaR and CVaR Calculations with Open FAIR™ Risk Analysis
Document No.: W241
Published by The Open Group, May 2024.
Any comments relating to the material contained in this document may be submitted to:
The Open Group, Apex Plaza, Forbury Road, Reading, Berkshire, RG1 1AX, United Kingdom
or by email to: ogspecs@opengroup.org

Table of Contents

Boundaryless Information Flow™

Boundaryless Information Flow™
achieved through global interoperability
in a secure, reliable, and timely manner

Executive Summary

In the White Paper *Calculating Reserves for Cyber Risk: Integrating Cyber Risk with Financial Risk* [W215], we showed how analysts using The Open Group Open FAIR™ Standards can measure cyber risk in economic terms as a distribution. From that distribution, Value at Risk (VaR) and Conditional Value at Risk (CVaR) calculations were made and integrated with other financial risks to render a holistic, economic risk picture against which reserves could be calculated. The example in that document demonstrated how Financial Institutions (FIs) can make Risk-Adjusted Return on Capital (RAROC) decisions and can comply with regulations that emphasize Tier 1 reserve management.

Effective VaR and CVaR measurements depend upon accurate models of the tail of each risk factor's distribution. The calibrated estimating technique, however, does not address the low frequency, high impact events analysts need to calculate VaR and CVaR. Without adjusting calibrated estimates to infer tail events, VaR and CVaR calculations derived from those estimates are meaningless. Tail events missing from the calibrated estimates can be inferred by extending a bounded distribution or by properly applying a calibrated estimate to specify an unbounded distribution. By adjusting a calibrated estimate in this way, analysts can make meaningful VaR and CVaR calculations to support accurate, compliant reserve calculations in FIs.

Expressing meaningful cyber risk estimates in the same terms as other financial risk estimates removes a historical boundary between cyber risk and other enterprise risk, making the information flow between enterprise risk managers and cyber risk managers boundaryless.

Introduction

This document is the third White Paper in the Calculating Reserves for Cyber Risk series.

In the Open FAIR™ Body of Knowledge – which comprises The Open Group Standard for Risk Taxonomy [O-RT] and The Open Group Standard for Risk Analysis [O-RA] (hereafter referred to as the Open FAIR Standards) – all risk factors are considered random variables. The Open FAIR Standards refer to and depend upon making calibrated estimates of the Minimum (Min), Maximum (Max), and Most Likely (ML) values to specify the range of a risk factor followed by a choice of that factor's Distribution (Dist). For example, the risk factors Loss Event Frequency (LEF) and Loss Magnitude (LM) may be specified as shown in Table 1.[1]

Table 1: Example Inputs for an Open FAIR Analysis

	LEF	**Loss Magnitude (LM)**			
		Primary Loss	**Secondary Loss**		
			SLEF[2]	**Secondary Loss Magnitude**	
	(Events/Yr)	Response ($)	Probability	Response ($)	Fines & Judgments ($)
MIN	0.2	30,000	0.2	15,000	1,000,000
ML	0.5	100,000	0.3	25,500	1,200,000
MAX	1	200,000	0.5	60,000	1,500,000
DIST	Triangular	Triangular for each LM Risk Factor			

Risk factor estimates are forward looking. That is, the actual value of the risk factor is estimated at the beginning of a period, to be observed later at that period's end. An estimate is considered accurate when the observed value lies within the previously estimated range nine times out of ten. Further, half of the observations not captured within the range should be below the Min and the other half above the Max.

Calibrated estimates are not designed to capture observations that have less than a 10% probability of occurring, and when used to specify bounded distribution choices, they deliberately exclude the 5% most impactful loss events from the analysis and risk model. It has been assumed that using calibrated estimates that ignore these outliers will not affect the decision support quality of the estimate.

But what if that assumption is not true? As discussed in first two White Papers in this Calculating Reserves for Cyber Risk series [W215, W221], the authors discussed the definition and importance of calculating Value at Risk (VaR) and Conditional Value at Risk (CVaR),[3] the significance that those calculations have on calculating reserves that comply with international banking law, and the role VaR and CVaR can have in

[1] These parameters are the same ones used in the example discussed in the White Paper *Calculating Reserves for Cyber Risk: Integrating Cyber Risk with Financial Risk* [W215].

[2] SLEF is short for Secondary Loss Event Frequency.

[3] See the White Paper *Calculating Reserves for Cyber Risk: Integrating Cyber Risk with Financial Risk* [W215] for full descriptions of VaR and CVaR.

making go/no-go investment decisions based upon RAROC measures.[4] VaR and CVaR look at low probability and high impact losses beyond a 95% confidence threshold. In the Calculating Reserves for Cyber Risk examples, those thresholds were at 99%.

Many cyber risk analysts use bounded distributions such as the uniform, triangular, or Beta-PERT distributions[5] to model risk factors and specify those distributions by a calibrated estimate. The tail information to accurately calculate VaR and CVaR calculations that is missing from those estimates raises many questions:

- What is the significance of calibrated estimates that only cover 90% of a risk factor's range, with 5% of the most significant losses (so-called "tail" events) deliberately not counted or estimated?
- How does the choice of distribution affect VaR and CVaR?
- How does an unbounded distribution, such as a log-normal distribution's VaR or CVaR, compare to bounded distributions' such as the uniform, triangular, or Beta-PERT?
- Can a calibrated estimate be adapted to infer the missing 10% of the range, and if so, how?
- How do calibrated estimates and choice of distribution affect VaR and CVaR calculations?
- How can analysts who need a VaR or CVaR use or extend calibrated estimating techniques to make reliable VaR and CVaR calculations?
- If calibrated estimates do not include expected outliers, and those outliers are important to supporting decisions, what is the Open FAIR analyst to do?

This document resolves some of these questions and gives some insight into the significance and implications of using calibrated estimating techniques to specify the minimum and maximum range of a distribution and the choice of the distribution itself. This document shows how to extend a calibrated estimate to infer the tail of a bounded distribution and how to apply calibrated estimation to unbounded distributions. Through these techniques, a calibrated estimate combined with a distribution can be used to make meaningful VaR and CVaR calculations.

[4] The examples used in the Calculating Reserves for Cyber Risk series are reused throughout this document. The terms and definitions used in those documents are used here as well, with the assumption that the reader knows and understands them.

[5] Note that the distribution types included in this document are examples. Any "true" distribution is likely unknown to the analyst, and analysts must recognize that any choice of distribution carries with it the uncertainty and potential model error associated with that choice being "wrong". Analysts must understand how that introduced uncertainty affects the decision quality of the decision the analysis is supporting. The distribution examples included in this document are illustrative of how modeled risk results might vary with a choice of distribution, but they do not indicate any preference for one distribution over any other in modeling cyber risk factors.

How Calibrated Estimates and Distribution Choices Influence Tail Risk Calculations: An Informatics Approach

The Open Group Standard for Risk Analysis (O-RA) [O-RA] discusses how a calibrated estimate characterized by a Min, Max, and ML value along with a Dist is the standard way to characterize Open FAIR risk factors as random variables. Tail risk – that is, VaR and CVaR – was not considered in the development of the O-RA Standard at that time. Subsequently, the first two White Papers in the Calculating Reserves for Cyber Risk series were published that shed light on the significance of tail risk in calculating reserves in an FI. What was not covered was a robust discussion of how tail risk can be evaluated from the combination of a calibrated estimate and choice of distribution. This section presents and elaborates upon two main points:

- The calibrated estimate by itself gives no insight into tail risk
- Tail risk depends upon the distribution chosen

Calibrated Estimates Alone Contribute Nothing to Understanding VaR

Cyber risk models that use calibrated estimate ranges alone omit all the data required to calculate VaR and CVaR. Consider the following example:

Suppose a calibrated estimate of ALE was between $10 and $100. By the definition of a calibrated estimate, five losses out of 100 would be below $10, and five losses out of 100 would be above $100.

Now suppose that the five losses above $100 are all at $150. That is, there is a 5% probability of a $150 loss, a 90% probability of a loss between $10 and $100, and a 5% probability of a loss at $5.

For CVaR calculation at the 95th percentile or beyond, the expected value in the tail is $150. That is to say that the calibrated estimate missed all relevant data informing a VaR or CVaR calculation. The calibrated estimate's range by itself offers no information whatsoever to the understanding of tail events at or beyond the 95th percentile.

While calibrated estimates of a single random variable account for 90% of all observations, the calibrated estimates of two random variables multiplied together, as is required to calculate risk from LEF and LM, account for only 81% of all observations, as indicated in Table 2.[6]

[6] Open FAIR models also add random variables together, such as summing two uniform random variables, which results in a triangular distribution. Tail analysis of the sum of random variables can be done similarly to this analysis.

Table 2: Observations that Contribute to a 99th Percentile Tail VaR and CVaR

<table>
<tr><td colspan="2"></td><td colspan="4">Loss Magnitude (LM)</td></tr>
<tr><td colspan="2"></td><td>0-5%</td><td colspan="2">Calibrated Estimate Range (5-95%)</td><td>95-100%</td></tr>
<tr><td rowspan="3">LEF</td><td>0-5%</td><td>0.25%</td><td colspan="2">4.5%</td><td>0.25%</td></tr>
<tr><td rowspan="2">Calibrated Estimate Range (5-95%)</td><td rowspan="2">4.5%</td><td colspan="2" rowspan="2">Data Used from Calibrated Estimate
81% of total data (0.9*0.9)</td><td>3.75-4.5%</td></tr>
<tr><td>0-0.75%</td></tr>
<tr><td>95-100%</td><td>0.25%</td><td>3.75-4.5%</td><td>0-0.75%</td><td>0.25%</td></tr>
</table>

In Table 2, the 99th percentile tail of the product LEF and LM is somewhere within the gray shaded area.[7] The exact location within that area depends upon the relative distribution differences between LEF and LM. Here again, unadjusted calibrated estimates cannot provide the information needed for VaR and CVaR analysis.

The Significance of Selecting a Distribution to Model a Risk Factor

Analysts and risk factor estimators will be uncertain about which distribution best models each cyber risk factor. However, choosing a single distribution to model each risk factor implies that the true distribution is known "with certainty". Typically, analysts do not know with certainty the closest-fit distribution for any cyber risk factor and are presenting results that claim more certainty than warranted by the available data and knowledge of the cyber risk landscape.

This section gives an example of how risk results change with the choice of distribution. Analysts can use this as an example of a sensitivity analysis of distribution choice and to justify developing stronger rationales behind their choices of distributions.

Distributions such as the uniform, triangular, or Beta-PERT are bounded by their minimum and maximum values. Analysts who use these distributions in Open FAIR analyses, as complete specifications of the range of those distributions, deliberately exclude 10% of all observations from the analysis. Of those excluded observations, half will be below the minimum and half above the maximum of the specified distribution.

Absent analysis, it is unclear what effect that exclusion has on risk results, particularly VaR and CVaR. It is also unclear how the selection of a bounded distribution affects those results. Table 3 shows the results for the risk scenario described in Table 1 using the Min, Max, and ML values directly entered in a Monte Carlo Simulation (MCS) risk modeling tool that can model uniform, triangular, or Beta-PERT distributions.

[7] Note the shaded area represents 1.75% of all possible observations. The tail's 1% observations lie somewhere within that 1.75%.

Table 3: Results of Using the Same Calibrated Estimates Across Three Bounded Distributions

Distribution Used	Average	ALE at 99th Percentile	VaR at 99th Percentile	CVaR at 99th Percentile
Uniform	333,370	2,843,659	2,509,959	2,906,663
Triangular	299,401	2,687,291	2,387,890	2,776,844
Beta-PERT	262,507	2,579,159	2,316,652	2,532,229

The uniform distribution with its comparatively fat tail has the highest 99th percentile Annual Loss Expectancy (ALE), VaR, and CVaR results. The Beta-PERT with its narrow tail bounded by the Max value has a very low probability of selecting values near the maximum end of the range, which explains its comparatively lower results. The triangular distribution lies somewhere in between.

Whether any of these differences matter depends upon their relevance to the question the analysis supports. If the difference between the numbers matters for the decision, then further investigation into the most relevant distribution is worth pursuing. If any choice of distribution suffices upon which to base the decision, then no further analysis is justified. In this example, if the uncertainty in VaR being between $2.317M and $2.524M (or CVaR between $2.60M and $2.95M) does not change a decision, then no further analysis is needed. Any choice of distribution would be sufficient to support that decision.

As discussed in the next section, calculating VaR and CVaR requires calibrated estimates to infer the tail of a bounded distribution.

Using Calibrated Estimates to Infer a Tail of a Bounded Distribution

How significant are the two missing tails of a calibrated estimate-specified bounded distribution? How can an analyst add to a 90% overall accurate range estimate the two missing 5% tails? This section extends the example of the three distributions used in Table 3 to reflect an inferred, but otherwise missing, left and right tail from distributions specified solely by an accurate, calibrated estimate.

Conceptually, adding an inferred tail to a distribution consists of expanding the distribution's minimum and maximum specification while preserving:

- The distribution's ML value, or, in the case of the uniform distribution, which has no ML value, its mean
- All other parameters that define the distribution outside of its minimum and maximum range, such as the shape parameter of a Beta-PERT distribution

The simplest example of inferring a tail on a distribution is the uniform distribution case. A 90% accurate calibrated estimate of a uniform distributed random variable can be extended by simply multiplying the calibrated estimate's total range (Min-Max) by 1.111 (that is, divided by 90%) and keeping the calibrated estimate's midpoint. Figure 1 shows the overall process of how a uniform distribution specified by an accurate calibrated estimate can be transformed into an inferred tail uniform distribution. The math is left up to the reader.

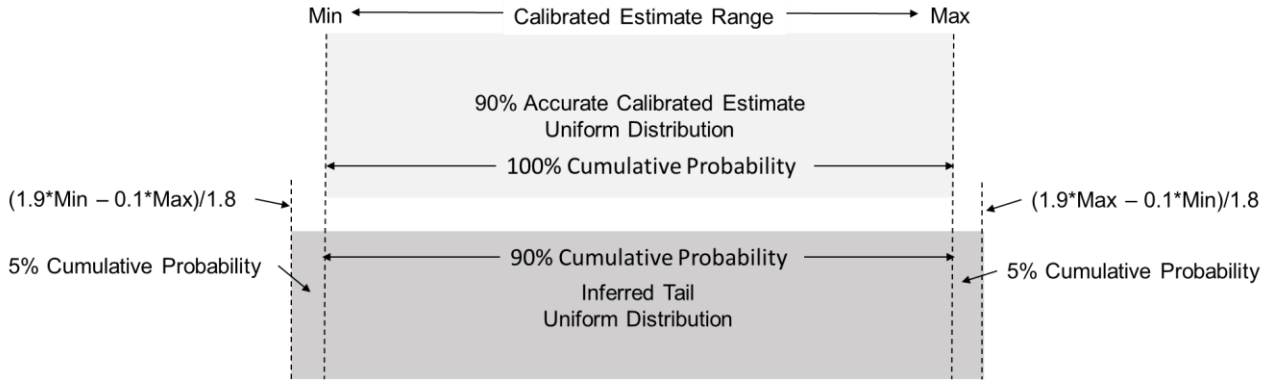


Figure 1: Inferring a Tail onto a Uniform Distribution

Inferring tails onto triangular and Beta-PERT distributions follows a similar process to that shown in Figure 1, while staying consistent with the concepts of preserving the ML and other specifications of the distribution. In each case,[8] only the distribution's minimum and maximum values are adjusted to infer (derive) the tail from the calibrated estimate.

In the examples that follow, the risk factors described in Table 1 are modeled as uniform, triangular, and Beta-PERT (shape parameter 4) distributions. All risk factors share the same bounded distribution. For

[8] That said, see the compromise made for the triangular distribution later in this document.

example, the LEF, SLEF, and each form of loss (Primary Response, Secondary Response, Fines and Judgments, etc.) are all modeled as either uniform, triangular, or Beta-PERT.

For each example below, the risk factor estimates shown in Table 4 are modeled with their respective distributions.

Table 4: Risk Factor Parameter Estimates Used Throughout This Document

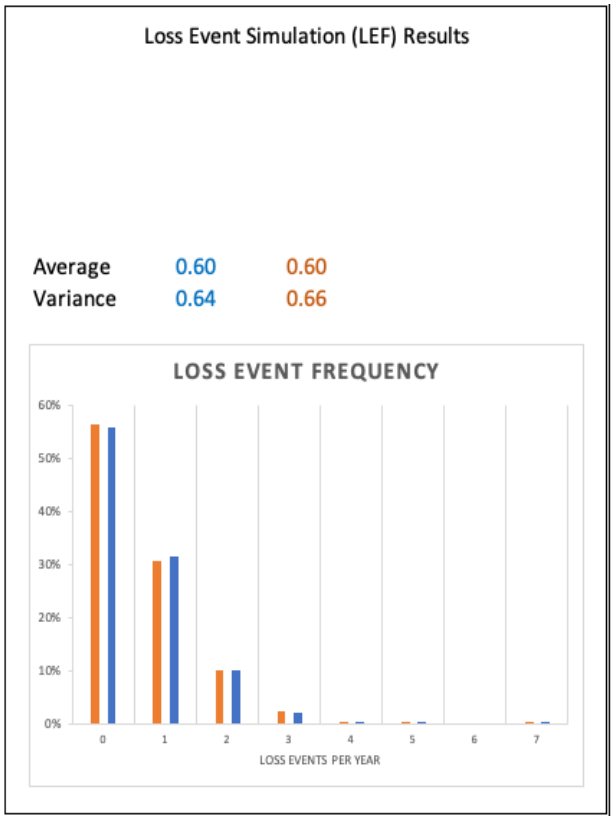
	LEF	**LOSS MAGNITUDE (LM)**			
		PRIMARY LOSS	**SECONDARY LOSS**		
			SLEF	**SECONDARY LM**	
	(Events/Yr)	Response ($)	Probability	Response ($)	Fines & Judgments ($)
MIN	0.2	30,000	0.2	15,000	1,000,000
ML	0.5	100,000	0.3	25,500	1,200,000
MAX	1	200,000	0.5	60,000	1,500,000

Example 1: Inferring a Tail on a Uniform Distribution

Modeling each risk factor shown in Table 4 as a uniform distributed random variable and then inferring tails on each of those random variables is analyzed below. The "current" state analysis is the uniform distribution as specified through the calibrated estimate. The "proposed" state analysis is the uniform distribution specified through the inferred tail process described in the previous section.

Inputs to the Uniform Distribution Model

The inputs to the model are shown in Figure 2 and Figure 3. Each distribution is uniform, and the proposed state shows how the specifications of the uniform distribution were modified to infer the left and right tail from the calibrated estimate specified in the current state.



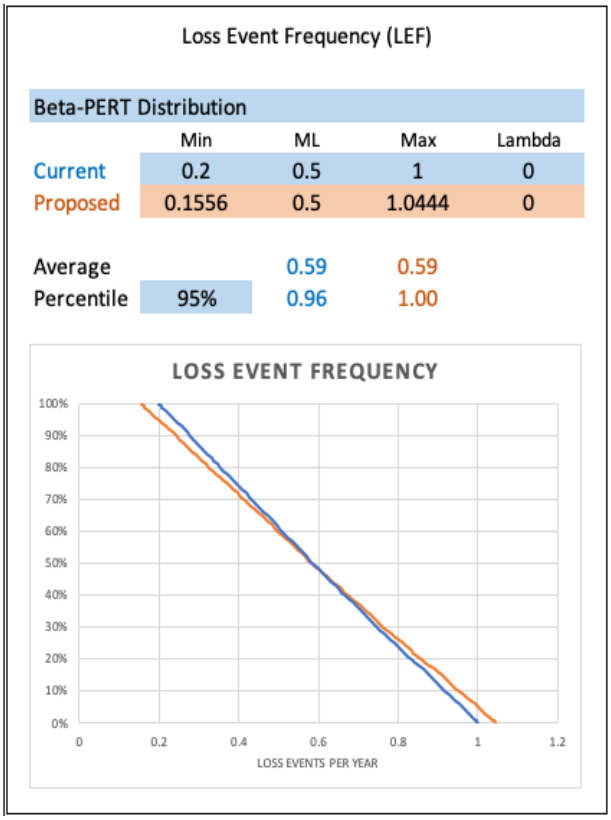


Figure 2: LEF Inputs to the Uniform Distribution Example[9]

[9] The uniform distribution is also modeled as a Beta-PERT distribution with shape factor (λ) equal to zero. The ML parameter has no meaning and should be ignored for this distribution.

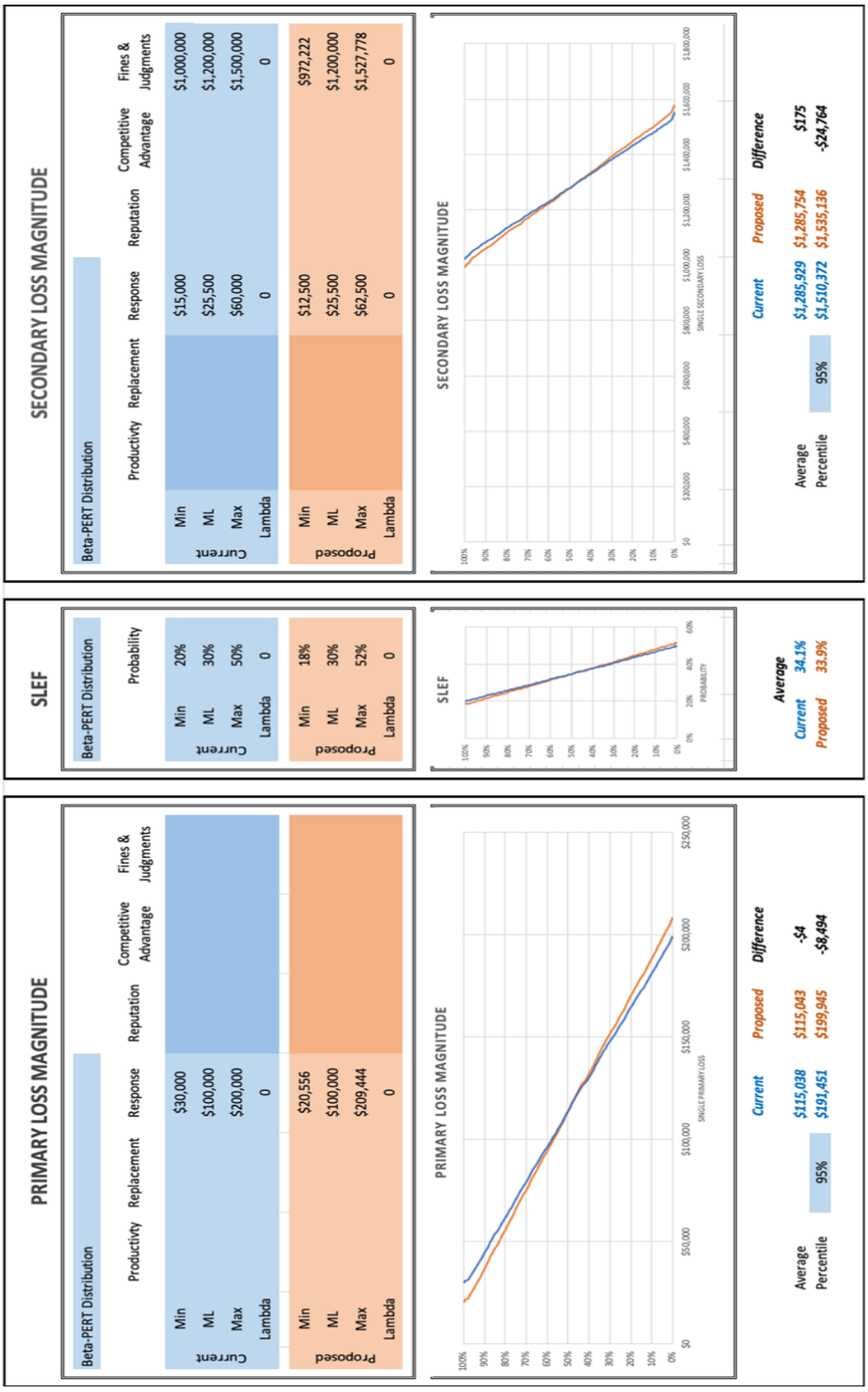


Figure 3: LM Inputs to the Uniform Distribution Example

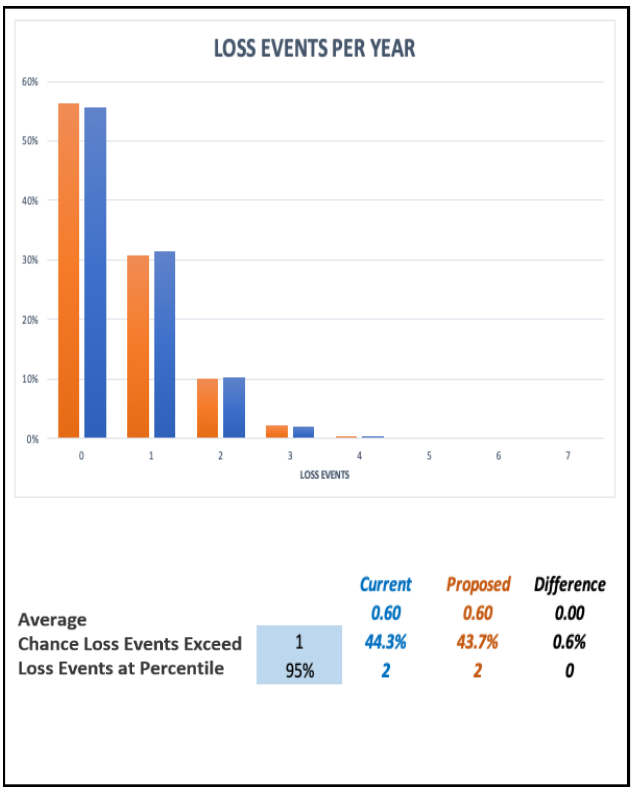
Table 5 shows how adjusting the endpoints of the uniform distribution to infer the implied missing information from an accurate calibrated estimate specification of the endpoints affects each risk factor's range. Note how these adjusted parameters vary from the original calibrated estimates. Analysts who are trying to capture these tails may not intuitively expect the percentage changes shown here. This is because a fixed offset is being added (subtracted) to the maximum (minimum) estimated value, then divided by the original maximum (minimum) value.

Table 5: Uniform Distributed Risk Factor Adjustments

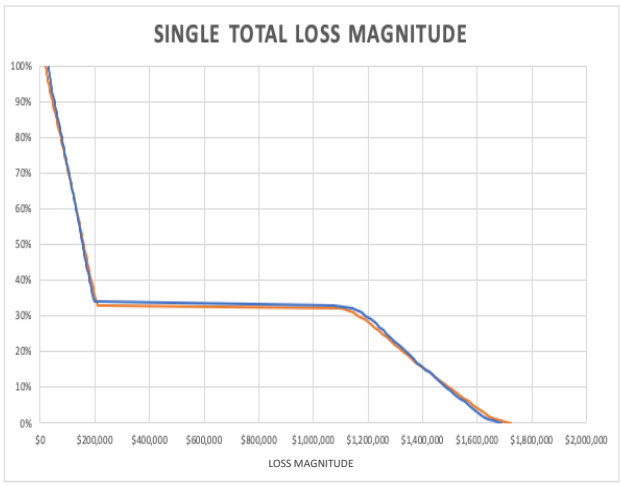
	Min Original	Revised	% Change	Max Original	Revised	% Change
LEF	0.2	0.1556	-22.2	1	1.044	4.4
Primary Response	30,000	20,556	-31.5	200,000	209,444	4.7
SLEF	20%	18%	-10	50%	52%	4.0
Secondary Response	15,000	12,500	-16.7	60,000	62,500	4.2
Secondary Fines & Judgments	1,000,000	890,616	-10.9	1,500,000	1,527,778	1.9

Results of Example 1: Uniform Distribution

Figure 4 and Figure 5 show the results of a 5,000-trial MCS of the calibrated estimate-specified uniform distributions for each risk factor (Blue) and the inferred tail specifications derived from the calibrated estimates (Orange).

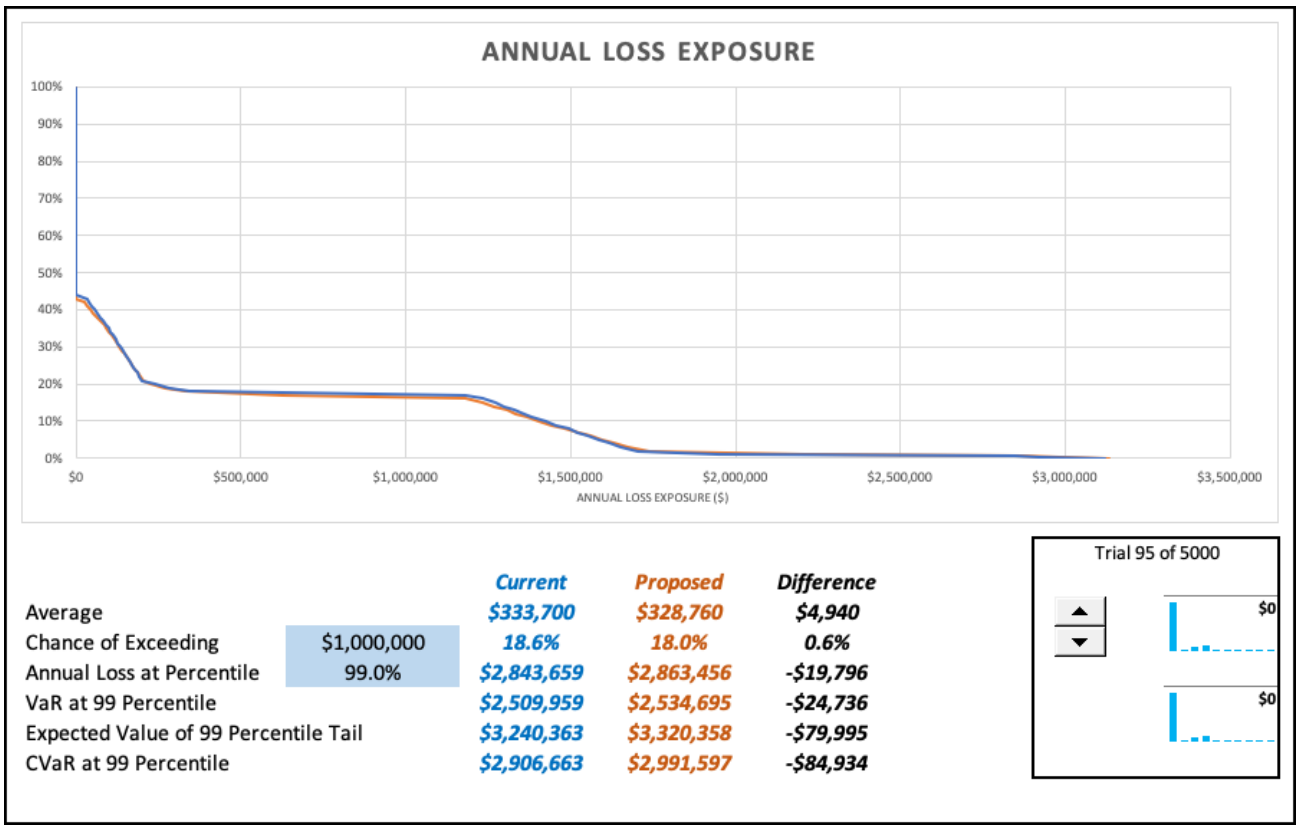


		Current	Proposed	Difference
Average		0.60	0.60	0.00
Chance Loss Events Exceed	1	44.3%	43.7%	0.6%
Loss Events at Percentile	95%	2	2	0



		Current	Proposed	Difference
Average		$552,940	$550,763	$2,177
Chance of Exceeding	$1,400,000	16.7%	16.5%	0.2%
Single Loss at Percentile	99.0%	$1,660,122	$1,687,519	-$27,398
VaR at 99 Percentile		$1,107,181	$1,136,756	-$29,575
CVaR at 99 Percentile		$1,134,944	$1,167,825	-$32,882

Figure 4: LEF and Single LM Results for the Uniform Distribution



		Current	Proposed	Difference
Average		$333,700	$328,760	$4,940
Chance of Exceeding	$1,000,000	18.6%	18.0%	0.6%
Annual Loss at Percentile	99.0%	$2,843,659	$2,863,456	-$19,796
VaR at 99 Percentile		$2,509,959	$2,534,695	-$24,736
Expected Value of 99 Percentile Tail		$3,240,363	$3,320,358	-$79,995
CVaR at 99 Percentile		$2,906,663	$2,991,597	-$84,934

Figure 5: ALE Results for the Uniform Distribution

It appears that while averages do not materially change, VaR and CVaR moderately do in this example. These results, however, may be in the expected statistical variation between individual MCS runs. If understanding the significance of these changes were necessary to support the decision the analysis supported, more simulations could be run and analyzed. In this example, it does not appear that inferring a tail on uniform distributions will make much of a difference. Other analyses, however, could have different results.

Example 2: Inferring a Tail on a Triangular Distribution

Figure 6 describes the goal of inferring a tail on a triangular distribution. The difficulty, however, is that simply extending the base of the triangle and holding the apex ML value of the triangle constant is incompatible with achieving a balanced left and right tail, each of which has a 5% cumulative probability. The only case where the left and right tails can be equal occurs when the apex is halfway in between the Min and Max values. That is, only when the triangular distribution is isosceles – that the ML is the average of the Min and Max calibrated estimate values – can the tail probabilities be equal.

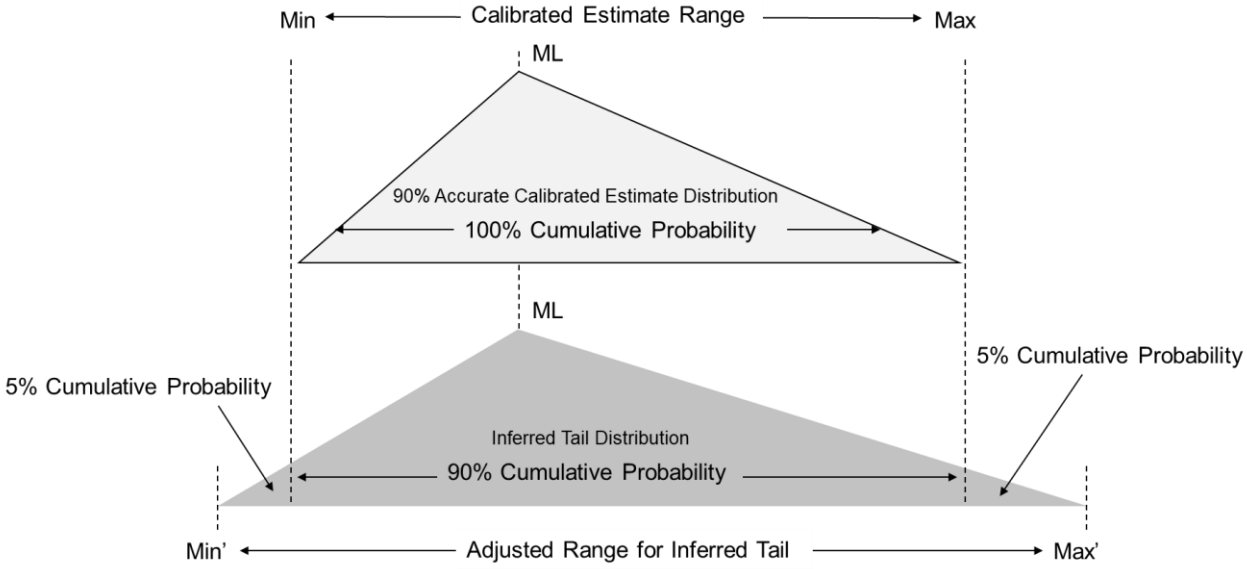
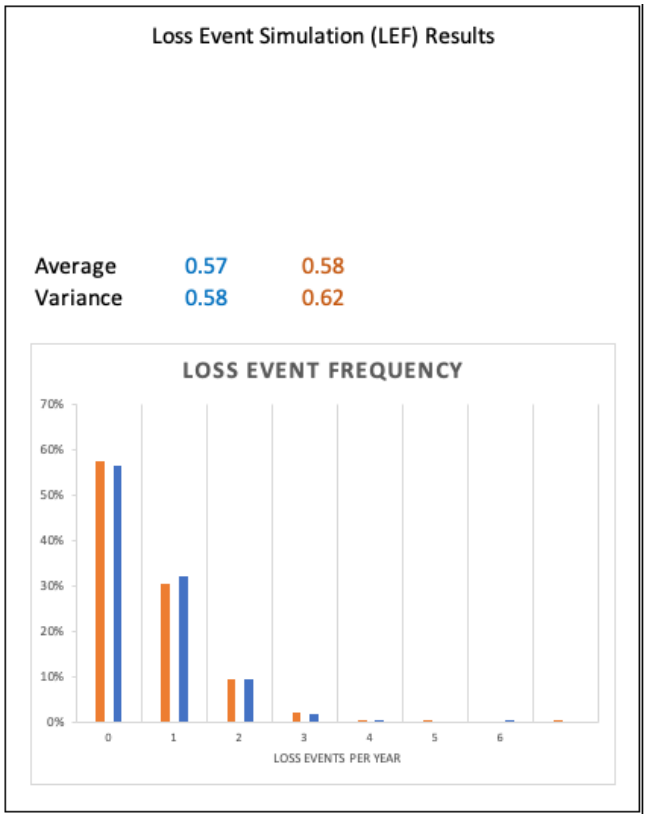


Figure 6: Goal of Inferring Tails on a Triangular Distribution

The analyst then must either relax the constraint of keeping the ML value the same or must accept that the left and right tails will not each be 5% cumulative probability, but hopefully something close. The choice made in this example for illustrative purposes was to maintain the ML value and accept the unequal spread of the total 10% cumulative probability between the left and right tail. Other analysts may make different decisions, especially when context specific to the analysis demands it.

Inputs to the Triangular Distribution Model

In the example that follows note that because each triangular distribution is skewed left, the adjusted (proposed) right-hand tail's 95th percentile values, as indicated in Figure 7 and Figure 8, are at or very close to the calibrated estimates' Max values – just what is needed for an accurate VaR and CVaR analysis.



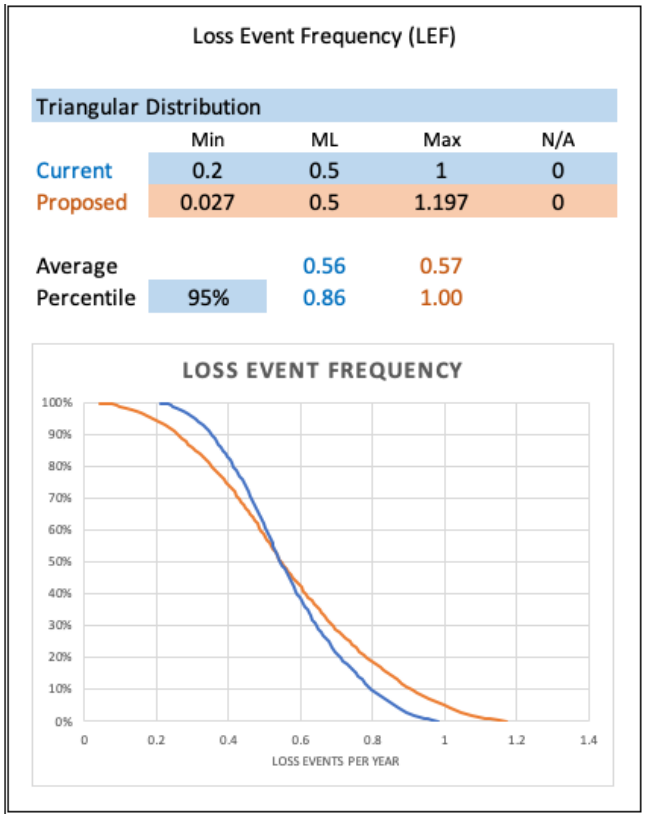


Figure 7: LEF Inputs to the Triangular Distribution

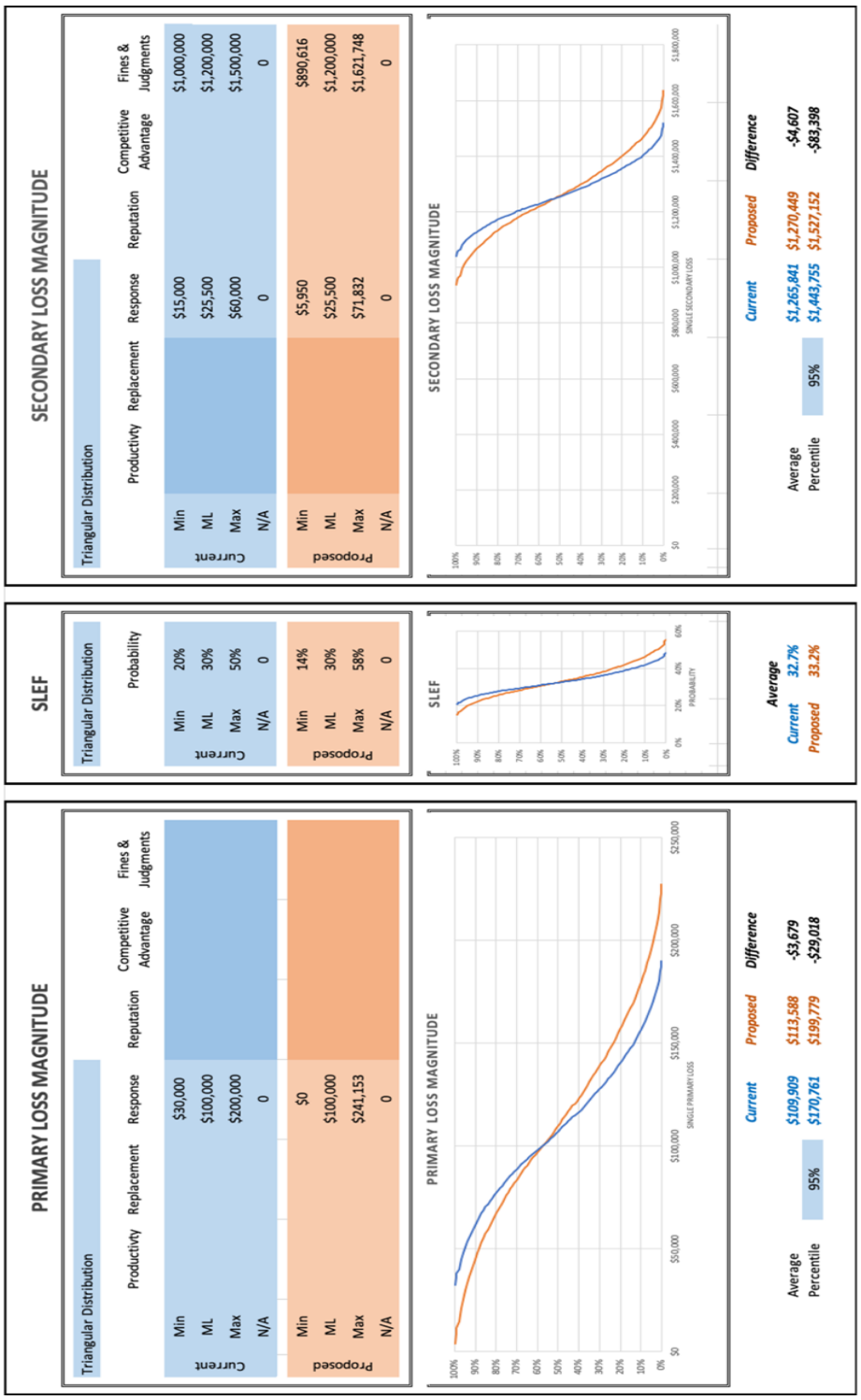


Figure 8: LM Inputs to the Triangular Distribution

Table 6 shows how the risk factor Min and Max values changed to infer the tail while maintaining the ML. To infer a 5% cumulative probability of observations below the calibrated estimate's minimum and a 5% cumulative probability of observations above the calibrated estimate's maximum requires adjustment of the risk factor estimates between -100 and +21%. The triangular distribution results are shown in Figure 9 and Figure 10. The magnitude of the adjustments needed may be unintuitive to the analyst who is trying to capture tail events from an assumed calibrated estimate. Additionally, the calculated Primary Response loss adjustment resulted in a minimum loss less than zero, something that violates the constraint of all loss estimates being greater than or equal to zero. The 0 value for the Min Revised is the pragmatic compromise made to stay within the constraint.

Table 6: Triangular Distributed Risk Factor Adjustments

	Min Original	Revised	% Change	Max Original	Revised	% Change
LEF	0.2	0.027	-86.5	1	1.197	+19.7
Primary Response	30,000	0	-100	200,000	241,153	+20.6
SLEF	20%	14%	-30.0	50	58	+16.0
Secondary Response	15,000	5,950	-60.3	60,000	71,832	+20.0
Secondary Fines & Judgments	1,000,000	890,616	-10.9	1,500,000	1,720,000	+14.7

Results of Example 2: Triangular Distribution

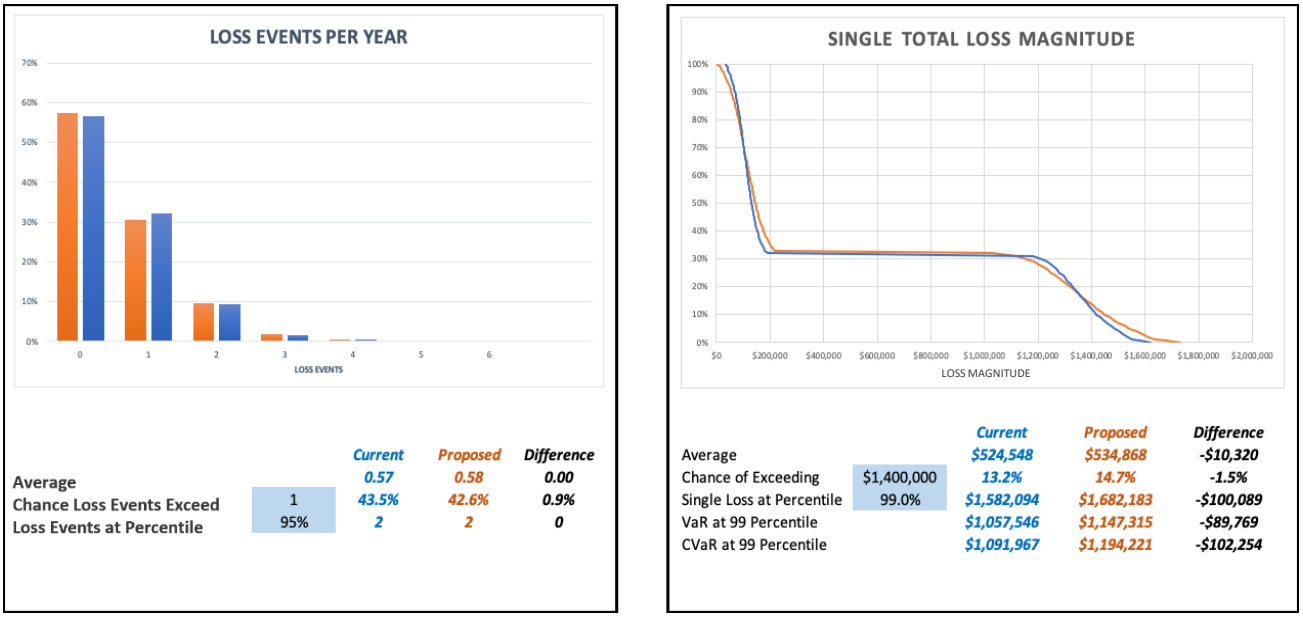


Figure 9: LEF and Single LM Results for the Triangular Distribution

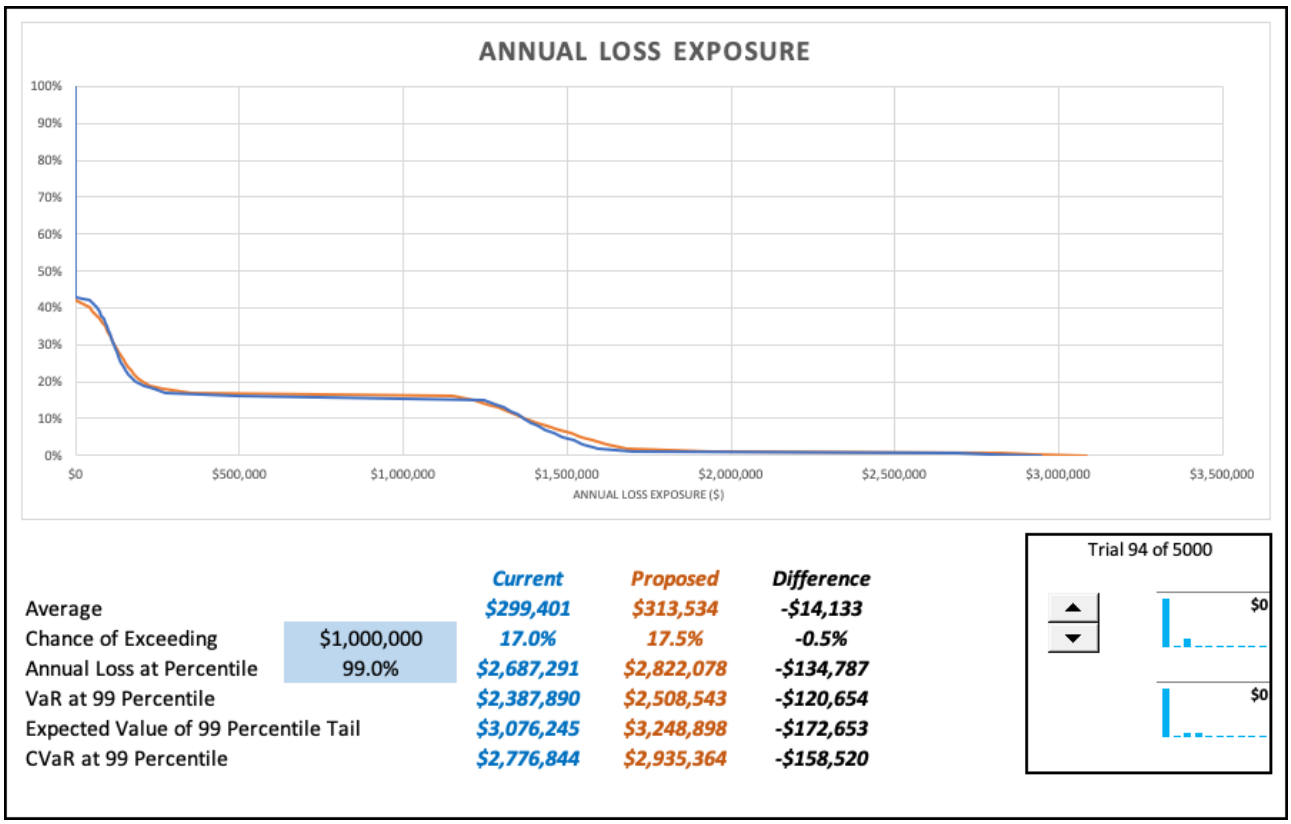
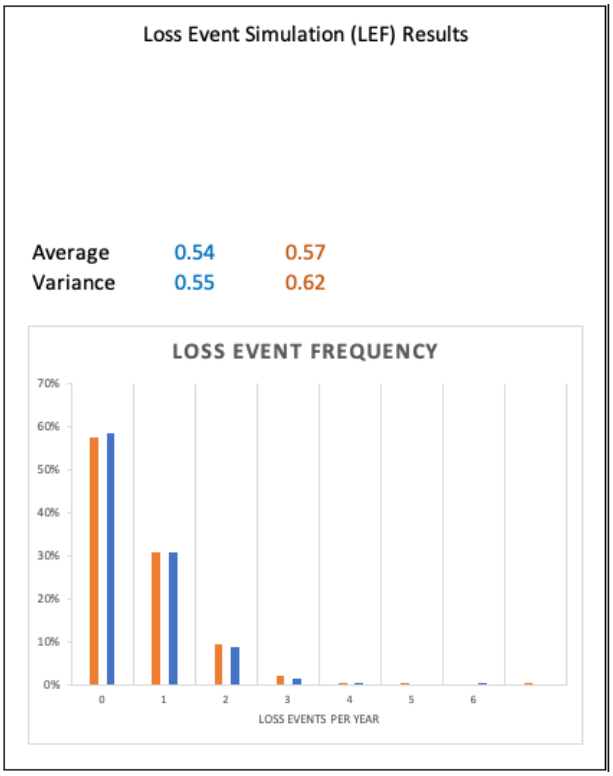


Figure 10: ALE Results for the Triangular Distribution

Example 3: Inferring a Tail on a Beta-PERT Distribution

Inputs to the Beta-PERT Analysis

Inputs for Example 3 are shown in Figure 11. Through trial and error, the author adjusted (in Orange) the Beta-PERT example to arrive at a Beta-PERT distribution, shown in Figure 12, that maintained its original ML and provided a cumulative 5% probability of observations below the original minimum and a 5% probability of observations above the original maximum.



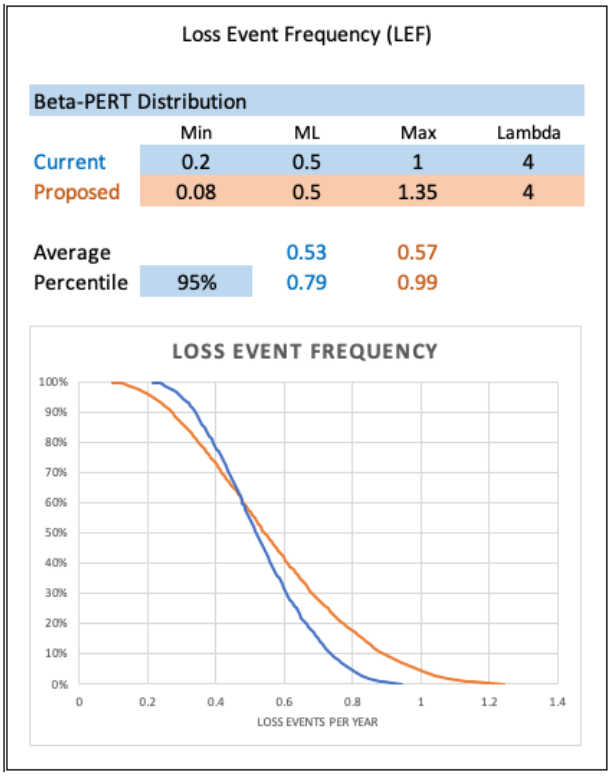


Figure 11: LEF Inputs to the Beta-PERT Distribution

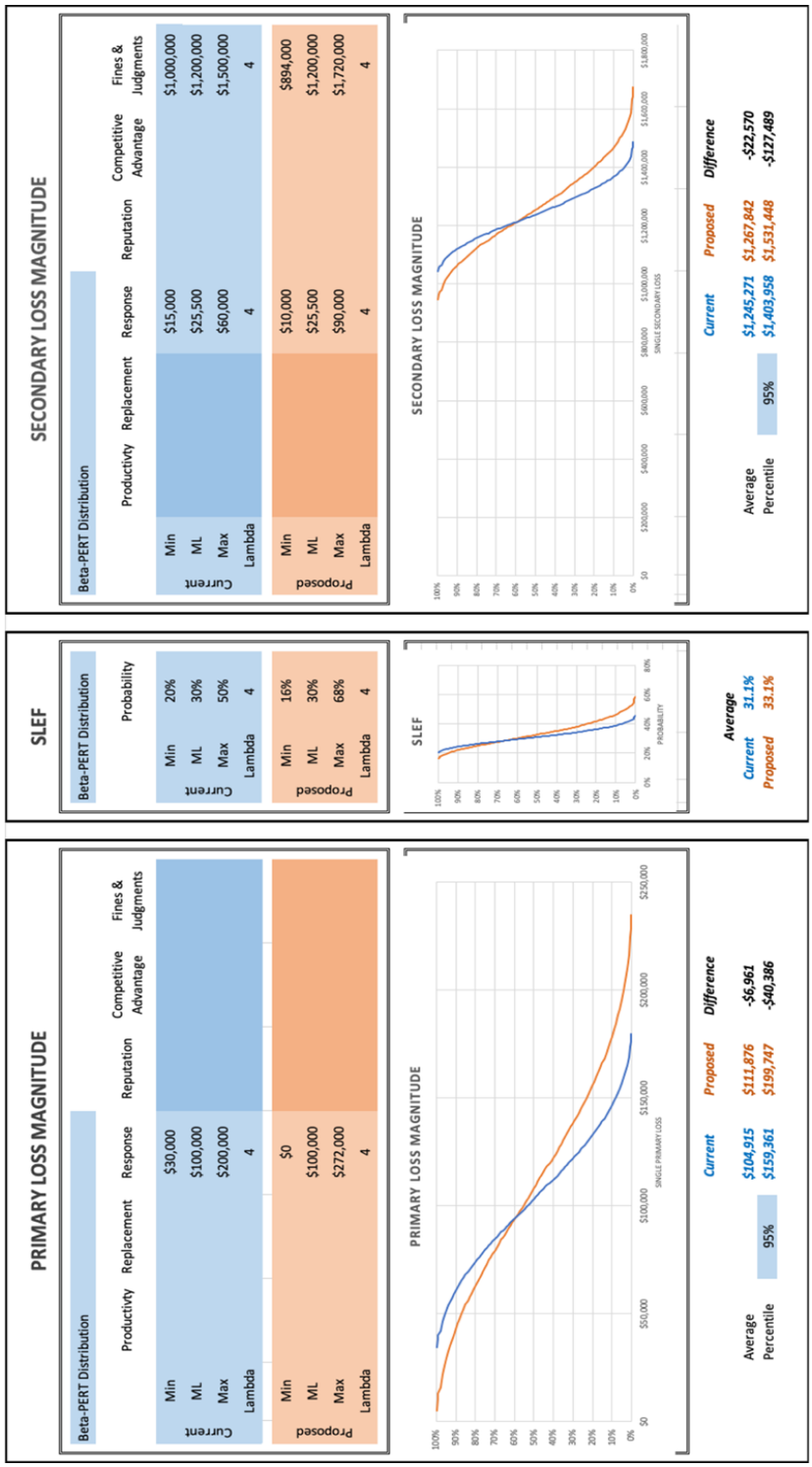


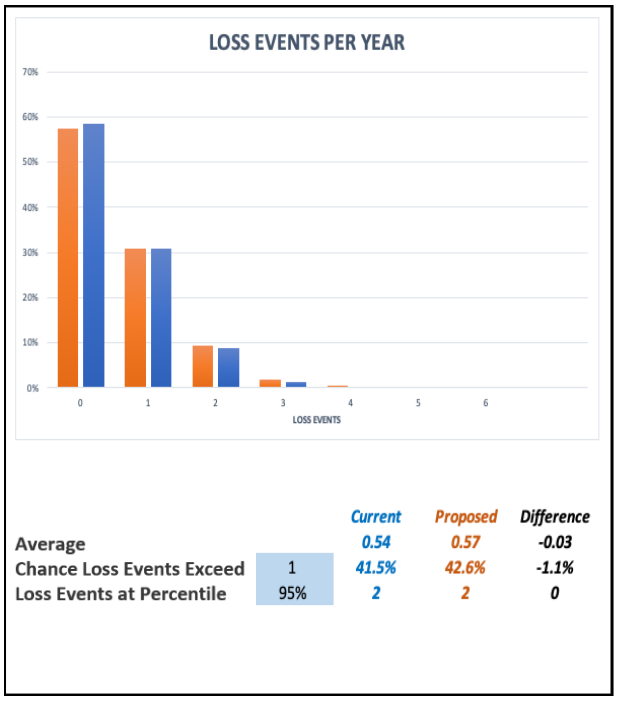
Figure 12: LM Inputs to the Beta-PERT Distribution

Table 7 shows how the risk factor Min and Max values changed to infer the tail while maintaining the ML. To infer a 5% cumulative probability of observations below the original minimum and a 5% cumulative probability of observations above the original maximum requires adjustments of the risk factor estimates between -100 and +50%. Results of the Beta-PERT Analysis for this example are shown in Figure 13 and Figure 14. This result may be unintuitive to the analyst who is trying to capture tail events from a calibrated estimate.

Table 7: Beta-PERT Distributed Risk Factor Adjustments

	Min Original	Revised	% Change	Max Original	Revised	% Change
LEF	0.2	0.08	-60	1	1.35	+35
Primary Response	30,000	0	-100	200,000	272,000	+36
SLEF	20%	16%	-20	50%	68%	+36
Secondary Response	15,000	10,000	-33	60,000	90,000	+50
Secondary Fines & Judgments	1,000,000	894,000	-11.6	1,500,000	1,720,000	+14.7

Results of the Beta-PERT Analysis



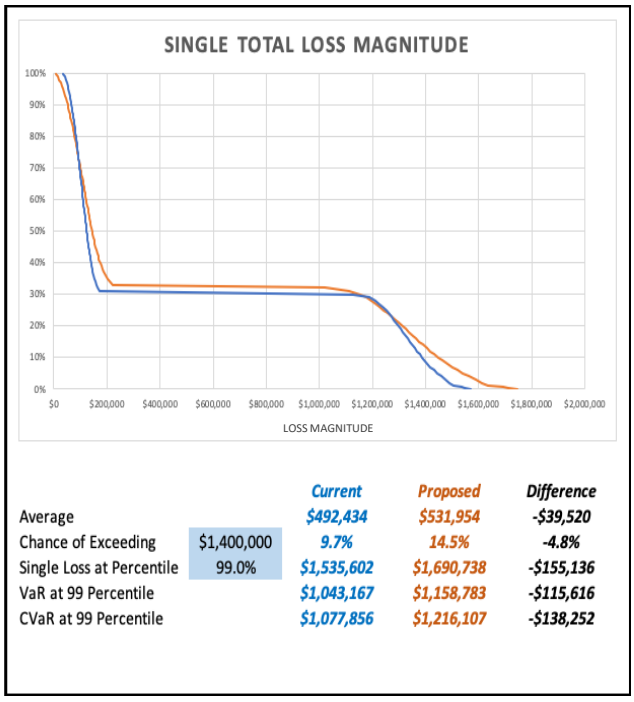


Figure 13: LEF and Single LM Results for the Beta-PERT Distribution

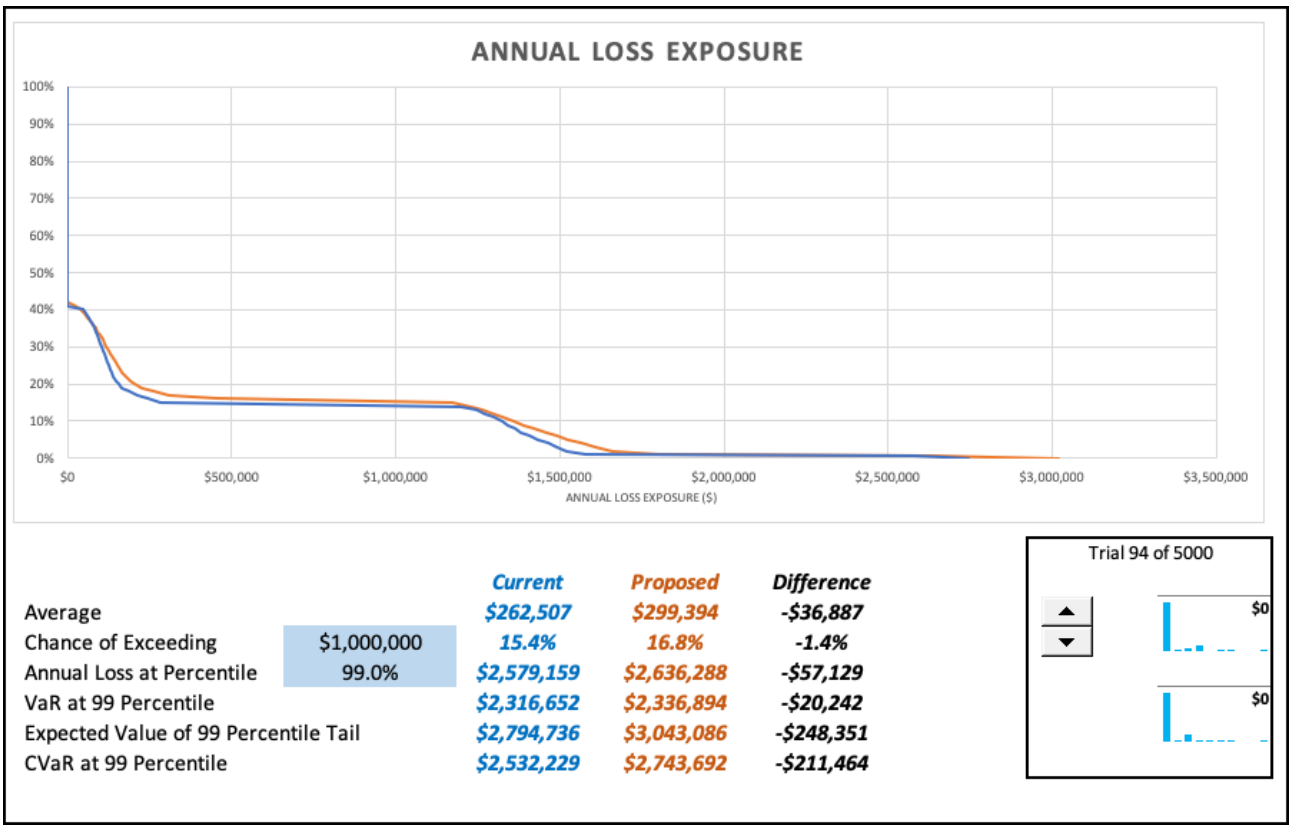


Figure 14: ALE Results for the Beta-PERT Distribution

Conclusions on Inferring a Tail on Bounded Distributions

The risk scenario in Table 1 was simulated under three scenarios, with each risk factor's random variable distribution modeled as:

- Uniform
- Triangular
- Beta-PERT

For simplicity, the same distribution was used to model each risk factor in each scenario. For example, in the uniform model, each risk factor LEF, Primary Loss Response Cost, SLEF, Secondary Response Cost, and Secondary Fines and Judgments were all modeled using the uniform distribution. No doubt that an analyst could choose different distributions for each of the different risk factors, but for this document only a single distribution was used across all of them. Table 8 gives the results of those inferred tail simulations for the three bounded distributions.

Table 8: Summary of Bounded Distribution Results

Distribution	Average	ALE at 99th Percentile	VaR at 99th Percentile	CVaR at 99th Percentile
Inferred Tail – Uniform	328,760	2,863,456	2,534,695	2,991,597
Inferred Tail – Triangular	313,534	2,822,078	2,508,543	2,935,364
Inferred Tail – Beta-PERT	299,394	2,636,288	2,336,894	2,743,692

There are some important observations here.

The uniform distribution's metrics hardly change between the inferred tail and the simple calibrated estimate model. That makes some intuitive sense as the extension to the uniform distribution is symmetrical, with an equal probability of lower-than-calibrated estimate values being drawn during the MCS compared to higher-than-calibrated estimate values being drawn.

The Inferred Tail - Triangular distribution's results differ from the originally specified triangular distributed example, most significantly the CVaR result. It seems plausible that the skew of the most significant triangularly distributed variables was a key influencing factor. Had the most significant variables been skewed differently, the results would likely differ.

The Inferred Tail – Beta-PERT distribution's results are lower than the other two.

Table 9 and Table 10 show the change in simulated risk outcomes due to inferring the tails on each of the bounded distributions. There are certainly differences, but whether they are relevant to decision-making rests upon the context of the decision and its criteria.

Table 9: Key Risk Measures Compared Between Standard Calibrated Estimates and Inferred Tail Estimates

Distribution	Average	ALE at 99th Percentile	VaR at 99th Percentile	CVaR at 99th Percentile
Uniform	332,135	2,857,126	2,524,990	2,951,815
Inferred Tail - Uniform	328,205	2,931,356	2,603,150	2,941,815
Triangular	296,828	2,697,825	2,400,997	2,637,551
Inferred Tail - Triangular	301,837	2,785,017	2,483,180	2,932,421
Beta-PERT	267,230	2,585,069	2,317,839	2,603,249
Inferred Tail – Beta-PERT	304,215	2,820,712	2,516,496	2,896,058

Table 10: Key Tail Result Percentage Change Due to Tail Inference

Distribution	Average	ALE at 99th Percentile	VaR at 99th Percentile	CVaR at 99th Percentile
Uniform	-1.5%	0.7%	1.0%	2.9%
Triangular	4.7%	5.0%	5.1%	5.7%
Beta-PERT	14.1%	2.2%	0.9%	8.4%

Using Calibrated Estimates to Infer Tails on Semi-Bounded or Unbounded Distributions

The previous sections discussed how tail events can be inferred when the distribution is bounded and the implications of not counting those tail events when measuring VaR and CVaR. This section discusses how to use accurate calibrated estimates to specify unbounded distributions, those that have no minimum or maximum values such as the normal distribution, or semi-bounded distributions, those that have one extreme (minimum or maximum) bounded and the other unbounded, such as the log-normal distribution.

Unbounded distributions that can be specified by a mean and Standard Deviation (SD), such as the normal distribution, usually have straightforward mathematical relationships between the 5th percentile (the Open FAIR minimum) and 95th percentile (the Open FAIR maximum) values. The ML may be flexible or constrained by the choice of the 5th percentile and 95th percentile values.

For example, as a Z Table shows, the 5th percentile value of a normal distribution is -1.645 SDs away from the mean. As the normal distribution is symmetric around its mean, the 95th percentile value is between +1.645 SDs away from the mean. The SD of a normal distribution characterized by an accurately calibrated estimate's minimum and maximum is approximately:

$$X \sim N(\mu, \sigma)$$

$$\sigma = (1.645 * 2)(Maximum - Minimum)$$

$$= 3.29 * (Maximum - Minimum)$$

With a mean μ of the distribution:

$$\mu = (Maximum + Minimum)/2$$

The normal distribution specified by mean and SD calculated this way when used in MCS will capture all tail events implied in the accurate calibrated estimate.

A word of caution here, though. Open FAIR risk factors are constrained in many ways that an unbounded distribution can violate. For example:

- Each form of loss (Productivity, Response, Replacement, Reputation, Competitive Advantage, and Fines and Judgments) and all frequencies (LEF, Threat Event Frequency, Contact Frequency) must be greater than or equal to zero
- All probabilities, such as the Probability of Action, Vulnerability, and SLEF must be between 0 and 1
- Threat Capability and Resistance Strength must be between 0 and 100

Unbounded distributions used to characterize Open FAIR risk factors will have to be adjusted or modified to reflect these limitations. In other words, they will have to have some bounds placed upon them.

A similar approach can be used to construct the parameters of the semi-bounded log-normal distribution. The lower limit of the distribution is 0 while the upper limit is infinite. As its name suggests, the random variable *x* is normally distributed, and the distribution itself is:

$$y = e^x \; ; \; \ln(y) = x$$

and:

$$x \sim N(\mu, \sigma)$$

That means that the natural logarithm of a 5th percentile accurate estimate of *y* is -1.645 SDs away from the natural logarithm of the exponent *x*'s mean, and the 95th percentile accurate estimate of *y* is between +1.645 SDs away from that *x*'s mean. Therefore, the SD of *x* is:

$$\sigma = (ln\ (maximum) - ln(minimum))/3.29$$

And the mean μ of *x* is:

$$\mu = (ln\ (maximum) + ln(minimum))/2$$

When used this way, an accurate calibrated estimate of a log-normal distribution can fully specify that distribution, including all information required to capture tail events.

A Semi-Bounded Distribution Example

Published reports suggest that losses appear to follow a log-normal distribution,[10] which provides the significance of using that distribution in this example to model each form of loss. The example that follows reuses the models of LEF and SLEF as Beta-PERT distributions; that is, the LEF and SLEF are duplicated as shown in the Beta-PERT example (see Figure 12). All forms of loss, however, are modeled here as log-normal distributions. Table 4 provides the estimates used for each risk factor.

This example only examines the effect of modeling the forms of loss as log-normal distributed losses instead of the bounded distributions used in the previous examples.

In this example, the Current (in Blue) LEF and SLEF inputs conformed to the unadjusted Beta-PERT specification based upon the calibrated estimate for LEF and SLEF in Table 4 and as replicated from Figure 12. The Proposed (in Orange) LEF and SLEF are inferred tail adjusted parameters used in the Beta-PERT example, as also shown in Figure 12.

[10] See IRIS 20/20 XTREME: Analyzing the 100 Largest Cyber Loss Events of the Last Five Years [IRIS 20/20].

Figure 15 reflects the Beta-PERT example inputs for LEF and SLEF. The example shows two scenarios:

- The Current scenario shows unadjusted LEF and SLEF Beta-PERT bounded distributions with all LM distributed log-normal

  The LEF and SLEF parameters used are the same as those in Figure 12.

- The Proposed scenario shows inferred tail adjusted LEF and SLEF Beta-PERT distributions with the same LM distributed log-normal as in the Current scenario

  The proposed scenario shows the effect of adjusting the LEF and SLEF parameters for the inferred tail to the Beta-PERT bounded distribution, as shown in Figure 12.

Inputs to the Semi-Bounded Log-Normal Distribution Example

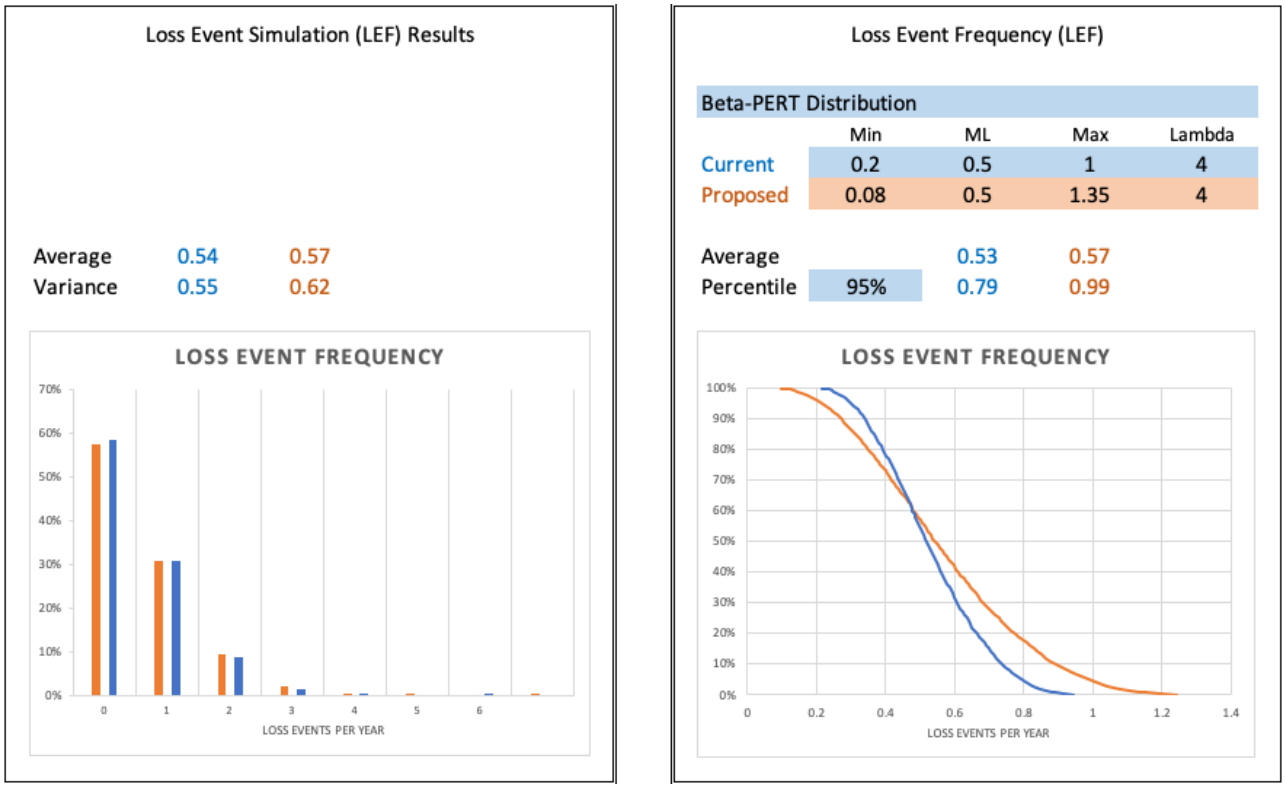


Figure 15: Beta-PERT Distributed LEF Inputs to the Log-Normal Distributed LM Risk Factor Example (Taken from Figure 12)

Figure 16 shows the results of modeling dollar-denominated losses as log-normal distributed.[11] Compare Figure 16 to Figure 12 to see the difference that the log-normal distribution of losses has on LM compared to a beta-pert distributed distribution of losses and the effect of adjusting the bounded LEF and SLEF distributions to infer tails on those bounds.

[11] Note that the ML estimates for the log-normal distributions are not relevant to any calculations and are ignored.

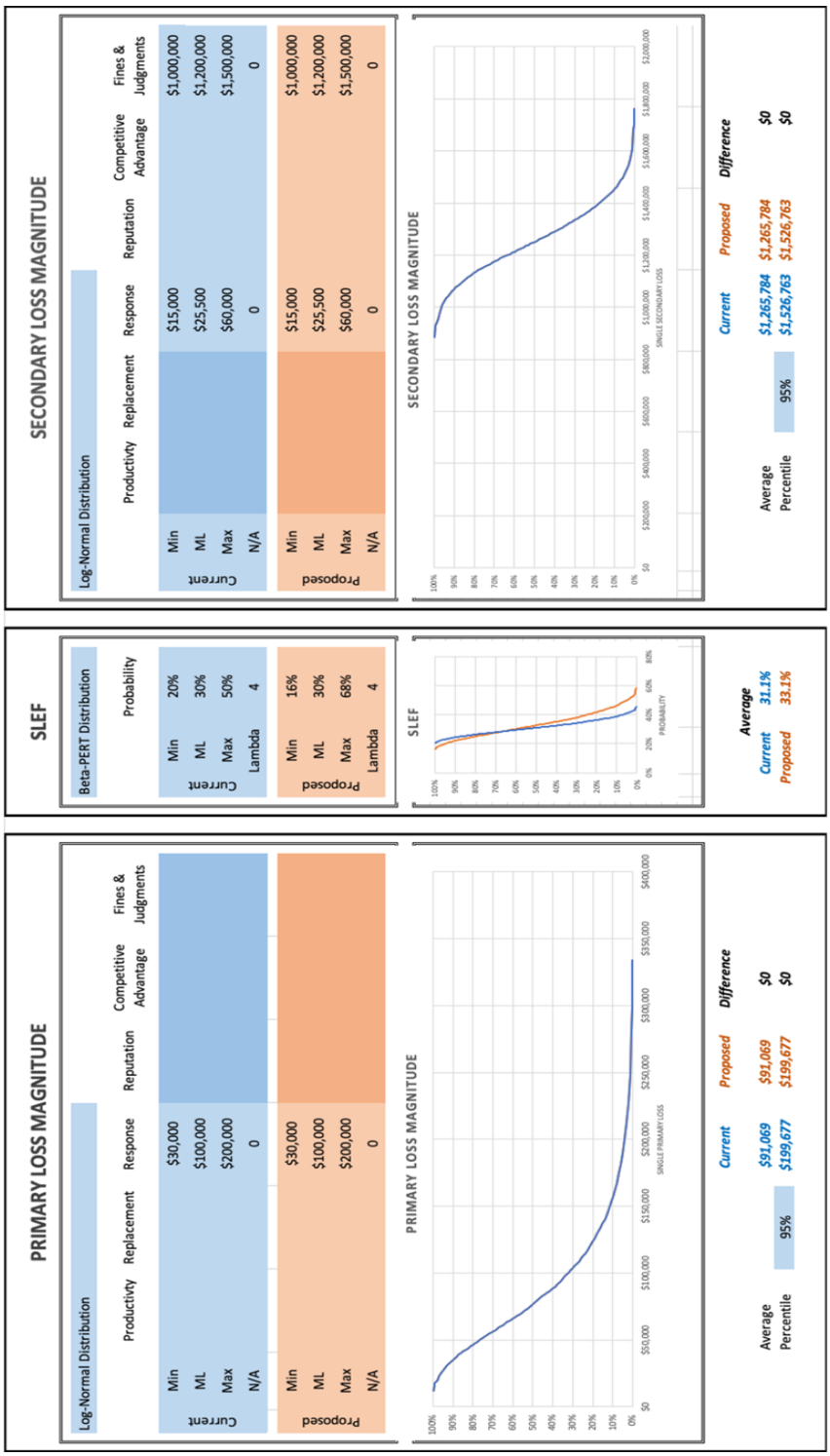


Figure 16: Log-Normal Distributed LM Risk Factors

Results of the Semi-Bounded Log-Normal Distribution Analysis

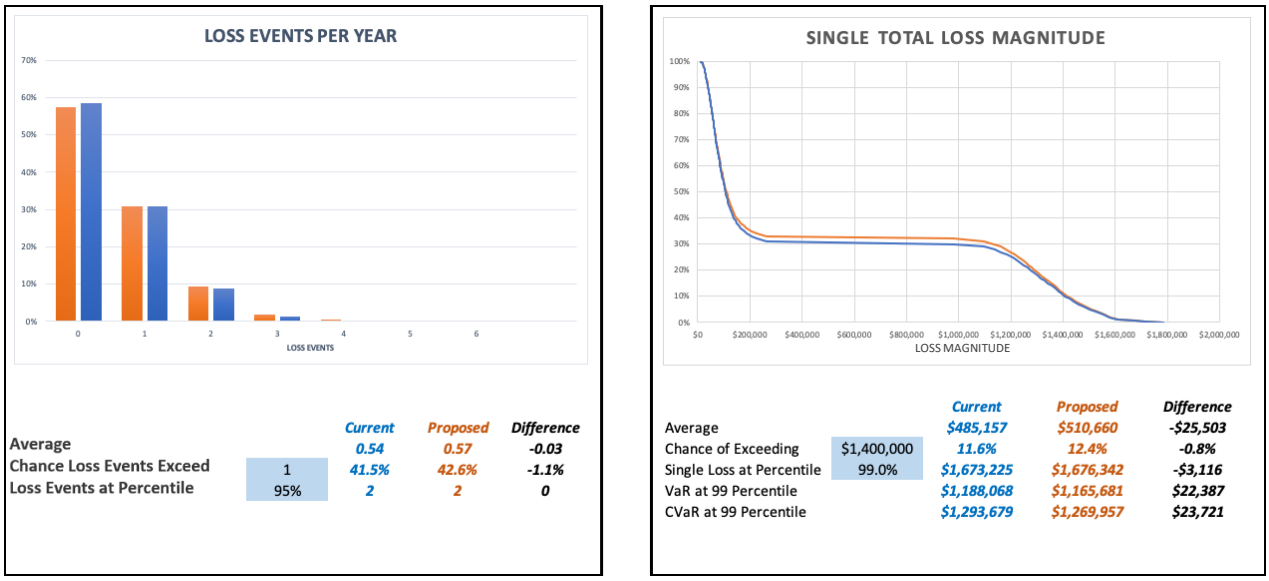


Figure 17: LEF and Single LM Results for the Beta-PERT Distributed LEF and SLEF and Log-Normal Distributed LM Risk Factors

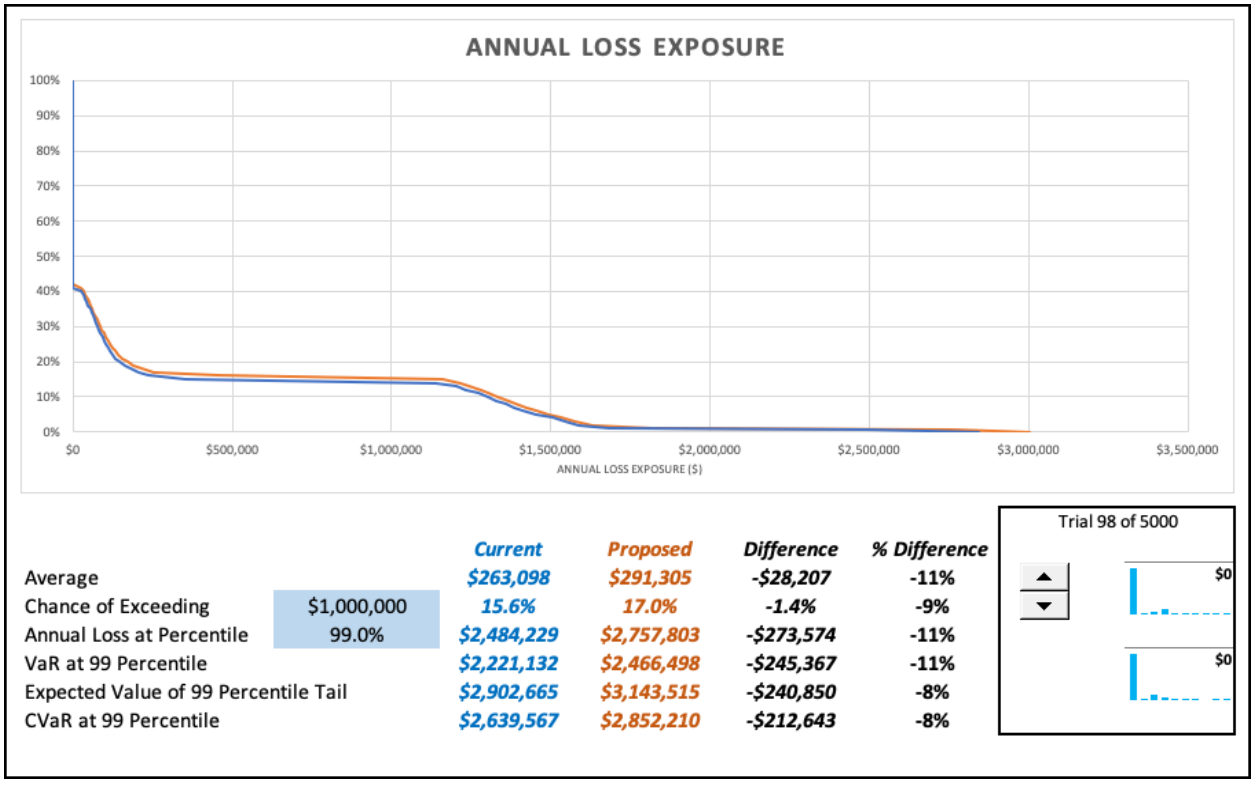


Figure 18: ALE Results Beta-PERT Distributed LEF and SLEF and Log-Normal Distributed LM Risk Factors

Figure 17 and Figure 18 show the difference that inferring the tail on LEF and SLEF has on results. The average and VaR both increased by 11% and CVaR increased by 8% due to inferring the tail on the two bounded distributions.

Table 11: Log-Normal Results Added to the Bounded Distribution Results

Distribution	Average	ALE at 99th Percentile	VaR at 99th Percentile	CVaR at 99th Percentile
Uniform	332,135	2,857,126	2,524,990	2,951,815
Inferred Tail – Uniform	328,205	2,931,356	2,603,150	2,941,815
Triangular	296,828	2,697,825	2,400,997	2,637,551
Inferred Tail – Triangular	301,837	2,785,017	2,483,180	2,932,421
Beta-PERT	267,230	2,585,069	2,317,839	2,603,249
Inferred Tail – Beta-PERT	304,215	2,820,712	2,516,496	2,896,058
Log-normal, Beta-PERT for LEF and SLEF	**263,098**	**2,484,229**	**2,221,132**	**2,639,567**
Log-normal, Inferred Tail Beta-PERT for LEF and SLEF	**291,305**	**2,757,803**	**2,466,498**	**2,852,210**

Table 11 shows how the log-normal example compares to all the previous analysis examples in this document. What surprised the author was how comparatively low the log-normal results were compared to the bounded distributions used to model losses. Many people intuitively think the log-normal distribution's infinite tail is comparatively fat, implying high VaR and CVaR implications. This example shows that may not be true.

That said, however, this document only examines one risk scenario. Analysts should not make any generalized conclusion from this single analysis.

Conclusion

This document explored the implications of ignoring tail events when making calibrated estimates to specify the range of distributions of risk factors. Specifically:

- What is the significance of calibrated estimates that only cover 90% of a risk factor's range, with 5% of the most significant events (so-called "tail" events) deliberately not counted or estimated?
- How does the choice of distribution affect VaR and CVaR?
- How does an unbounded distribution, such as a log-normal distribution's VaR or CVaR, compare to bounded distributions', such as the uniform, triangular, or Beta-PERT?
- Can a calibrated estimate be adapted to infer the missing 10% of the range, and if so, how?
- How do calibrated estimates and choice of distribution affect VaR and CVaR calculations?
- How can analysts who need a VaR or CVaR use or extend calibrated estimating techniques to make reliable VaR and CVaR calculations?
- If typical calibrated estimates do not include expected outliers, and those outliers are important to supporting decisions, what is the Open FAIR analyst to do?

This document has shown that analysts can infer the likelihood and magnitude of tail events from the combination of an accurate calibrated estimate and a statistical distribution. However, an important assumption is made here, that the calibrated estimate is accurate. That is, the estimate's lower bound is the 5th percentile of the distribution, and its upper bound is the 95th percentile. Estimators can be trained to generally meet this assumption. The analyst in each case needs to verify that assumption.

The distribution choices have some effect, but for many analyses – and the decisions they support – the uncertainty of what distribution best models any particular risk factor is less important than other considerations. Whether the analyst should obsess over what distribution to use to model each risk factor depends upon the sensitivity of results to the choice of distribution and to how that change in results would affect business decisions based upon the analysis. As with any model simulation and use of statistics, context and purpose matter.

When the decision requires it, analysts can vary the distribution used to model risk factors to see how sensitive the results are based on choice of distribution. That way, analysts can decide whether further research is worth doing to assess distribution best models the risk factor at hand. Usually, running analyses under various conditions is inexpensive, so empirical sensitivity analysis can be very feasible.

Accurate calibrated estimates along with the chosen distribution have all the information analysts need to model tail events. The key, however, is that the calibrated estimate is indeed accurate. A calibrated estimate is defined as "accurate" when its range includes 90% of all possible observations, with a 5% probability that an observation will be below the estimated range, and a 5% probability that an observation will be above the estimated range.

This document has shown how to take an accurate estimate along with a distribution to create parameters that respect the 90% overall accuracy of the estimate and the inferred tail events not specified within the estimate's range.

Extending the calibrated estimate can be done mathematically or numerically through trial and error. However, when no distribution exists that can maintain the ML and the 5% tail probabilities of observations lying outside the estimate, a compromise must be struck. The non-isosceles triangular distribution is an example where such a compromise must be accepted. Whether that compromise is material depends upon the context and purpose of the analysis.

Calibrated estimates can fully specify unbounded or semi-bounded distributions, such as the normal and log-normal distributions. Sometimes, however, these distributions will have to be modified so that the risk factors they are modeling conform to their definitional constraints. For example, an unmodified normal distribution, whose minimum value stretches infinitely negative, used to model a loss would violate the constraint that losses must always be positive. Similarly, a normal distribution unconstrained to be between 0 and 1 would be unsuitable for modeling any risk factor measured in probability. While these problems can be worked around, they cannot be ignored.

Using these techniques, risk analysts can use accurate calibrated estimates and distributions to model risk factors that include inferred extreme tail events. Of all things to consider, the calibrated estimate's accuracy, as the Open FAIR Standards define accuracy, is relied upon to indeed be accurate. Assessing the quality of the calibrated estimating process and choice of risk factor distributions must be part of any cyber risk model vetting process.

Improving the accuracy of cyber risk factor estimates will require substantially more research. In particular, more research is needed in:

- Integrating data with the calibrated estimation process

  For VaR and CVaR calculations to be accurate, the 5^{th} and 95^{th} percentile estimates along with the choice of distribution must be accurate. Although more data is being collected and communicated on the

frequency and magnitude of cyber losses, integrating that data and subject matter expertise to arrive at accurate estimates of future losses is still an area that needs more research.

- The distributions that most accurately model risk factors

  More research could argue for better choices of distributions. Failing that, more research could answer important questions on the nature of cyber risk factor distributions. One question that should be asked is whether risk factor distributions are coherent. If not, then VaR and CVaR calculations can break down.

- The calibrated estimating techniques incorporated into the O-RA Standard that are based upon how people with limited information estimate fixed attributes such as the height of the Sears Tower or the wingspan of a Boeing™ 747

  Whether those techniques can be directly applied to estimating future outcomes such as the cyber risk random variables LEF and LM has been assumed to be intellectually correct, but is that assumption true? More research on this topic is needed to reinforce the techniques developed in this document.

- All assumptions used in cyber risk analyses must be valid

  In particular, the Open FAIR Standards by default characterize each risk factor as an independent, identically distributed random variable. Is this assumption true? Are there correlations between risk factors that make them dependent upon each other? Is each loss event and LM independent of each other? To better model cyber risk, these questions and others related to the assumptions should be researched.

Finally, the techniques presented here can be used in any context, not just in the measurement of cyber risk in FIs, where accurate estimates of uncertain outcomes are based primarily upon human judgment that is informed by comparatively little data. Risk managers in many industries are beginning to see that they need all risks to be measured in economic terms so that multiple risks can be integrated, with tail events sufficiently and accurately estimated to support VaR, CVaR, and RAROC analysis, even for those risk events that are plausible but have never been observed. The techniques in this document provide ways forward to meet these emerging requirements.

Glossary

Action

An act taken against an Asset by a Threat Agent. Requires first that contact occurs between the Asset and Threat Agent.

Asset

The information, information system, or information system component that is breached or impaired by the Threat Agent in a manner whereby its value is diminished, or the act introduces liability to the Primary Stakeholder.

Banking Book

Refers to the activities in a bank, excluding the trading floor activities in a bank.

Competitive Advantage Loss or Cost

One of the six forms of loss, Competitive Advantage Losses are losses associated with diminished competitive advantage. Competitive Advantage Loss is specifically associated with Assets that provide competitive differentiation between the organization and its competition. Examples include trade secrets, merger and acquisition plans, etc.

Conditional Value at Risk (CVaR)

Conditional VaR is an alternate risk measure that gives an indication of the magnitude of the potential losses in the tail. In particular, CVaR is the expected loss beyond VaR (i.e., the expected loss given that the loss exceeds the VaR).

Contact Event

Occurs when a Threat Agent establishes a physical or virtual (e.g., network) connection to an Asset.

Contact Frequency (CF)

The probable frequency, within a given timeframe, that a Threat Agent will come into contact with an Asset.

Control

Any person, policy, process, or technology that has the potential to reduce the Loss Event Frequency – Loss Prevention Controls – and/or Loss Magnitude – Loss Mitigation Controls.

Control Strength (CS)

The strength of a control as compared to a standard measure of force.

FAIR

Factor Analysis of Information Risk.

Fines and Judgments Losses or Costs

One of the six forms of loss, Fines and Judgments losses or costs, are those associated with legal or regulatory actions levied against an organization. Note that this includes bail for any organization members who are arrested.

Loss Event

Occurs when a Threat Agent's action (Threat Event) is successful in breaching or impairing an Asset.

Loss Event Frequency (LEF)

The probable frequency, within a given timeframe, that a Threat Agent will inflict harm upon an Asset.

Loss Flow

The structured decomposition of how losses materialize when a Loss Event occurs.

Loss Magnitude (LM)

The probable magnitude of loss resulting from a Loss Event. Losses are categorized into six forms of loss: Productivity, Replacement, Response, Competitive Advantage, Reputation, and Fines and Judgments.

Loss Scenario

The story of loss that forms a sentence from the perspective of the Primary Stakeholder.

Operational Risk

Regulators define Operational Risk as "the risk of loss resulting from inadequate or failed internal processes, people, and systems or from external events". The Basel III Capital Accords consider seven Level 1 Loss Event types. Cyber risk is an operational risk under Basel III and can influence several of these Level 1 Loss Event types.

Parsimony

The goal that a risk model is neither too complex nor too simple.

Primary Stakeholder

The person or organization that owns or is accountable for an Asset.

Probability of Action (PoA)

The probability that a Threat Agent will act against an Asset once contact occurs.

Productivity Loss or Cost

One of the six forms of loss, Productivity Losses are losses associated with the reduction in an organization's ability to generate its primary value proposition (e.g., income, goods, services).

Regulatory Capital

Regulatory Capital is the minimum amount of capital imposed by the regulator.

Replacement Loss or Cost

One of the six forms of loss, Replacement Losses are those associated with the intrinsic value of an Asset. Typically represented as the capital expense associated with replacing lost or damaged Assets (e.g., rebuilding a facility, purchasing a replacement laptop).

Reputation Loss or Cost

One of the six forms of loss, Reputation Losses are those associated with an external perception that an organization's value proposition is reduced, or leadership is incompetent, criminal, or unethical.

Resistance Strength (RS)

The strength of a Control as compared to the probable level of force (as embodied by the time, resources, and technological capability; measured as a percentile) that a Threat Agent is capable of applying against an Asset.

Response Loss or Cost

One of the six forms of loss, Response Losses are the expenses associated with managing a Loss Event (e.g., internal or external person-hours, logistical expenses).

Risk

The probable frequency and probable magnitude of future loss.

Risk Analysis

The process to comprehend the nature of risk and determine the level of risk. [Source: ISO Guide 73:2009]

Risk Assessment

The overall process of risk identification, risk analysis, and risk evaluation. [Source: ISO Guide 73:2009]

Risk Factors

The individual components that determine risk, including Loss Event Frequency, Loss Magnitude, Threat Event Frequency, etc.

Risk Management

Coordinated activities to direct and control an organization with regard to risk. [Source: ISO Guide 73:2009]

Secondary Stakeholder

Individuals or organizations that may be affected by events that occur to Assets outside of their control. For example, consumers are Secondary Stakeholders in a scenario where their personal private information may be inappropriately disclosed or stolen.

Threat

Anything that is capable of acting in a manner resulting in harm to an Asset and/or organization; for example, acts of God (weather, geological events, etc.), malicious actors, errors, failures.

Threat Agent

Any agent (e.g., object, substance, human, etc.) that is capable of acting against an Asset in a manner that can result in harm.

Threat Capability (TCap)

The probable level of force (as embodied by the time, resources, and technological capability) that a Threat Agent is capable of applying against an Asset.

Threat Community

A subset of the overall Threat Agent population that shares key characteristics.

Threat Event

Occurs when a Threat Agent acts against an Asset.

Threat Event Frequency (TEF)

The probable frequency, within a given timeframe, that a Threat Agent will act against an Asset.

Trading Book

Refers to the trading activities in a bank.

Value at Risk (VaR)

Value at Risk is defined as the worst loss that might be expected from holding a security or a portfolio over a given period of time (say a single day or 10 days) given a specified level of probability known as the confidence level.

Vulnerability (Vuln)

The probability that a Threat Event will become a Loss Event; the probability that Threat Capability is greater than Resistance Strength. (Synonym: Susceptibility)

Acronyms & Abbreviations

ALE	Annual Loss Expectancy
CVaR	Conditional Value at Risk
Dist	Distribution
FAIR	Factor Analysis of Information Risk
FI	Financial Institution
LEF	Loss Event Frequency
LM	Loss Magnitude
Max	Maximum
MCS	Monte Carlo Simulation
Min	Minimum
ML	Most Likely
O-RA	The Open Group Standard for Risk Analysis
O-RT	The Open Group Standard for Risk Taxonomy
PERT	Program Evaluation and Review Technique
RAROC	Risk-Adjusted Return on Capital
SD	Standard Deviation
SLEF	Secondary Loss Event Frequency
VaR	Value at Risk

References

(Please note that the links below are good at the time of writing but cannot be guaranteed for the future.)

Basel III — Basel III Capital Accords; refer to: https://en.wikipedia.org/wiki/Basel_III

Hubbard 2014 — How to Measure Anything: Finding the Value of "Intangibles" in Business, Third Edition, Douglas W. Hubbard, April 2014, Wiley

IRIS 20/20 — Information Risk Insights Study (IRIS) 20/20 XTREME: Analyzing the 100 Largest Cyber Loss Events of the Last Five Years; published by Cyentia Institute, 2020; refer to: https://www.cyentia.com/wp-content/uploads/IRIS2020-Xtreme.pdf

ISO 73:2009 — ISO Guide 73:2009, Risk Management – Vocabulary, November 2009; refer to: https://www.iso.org/standard/44651.html

O-RA — The Open Group Standard for Risk Analysis (O-RA), Version 2.0.1 (C20A), published by The Open Group, November 2020; refer to: www.opengroup.org/library/c20a

O-RT — The Open Group Standard for Risk Taxonomy (O-RT), Version 3.0.1 (C20B), published by The Open Group, November 2020; refer to: www.opengroup.org/library/c20b

W215 — Calculating Reserves for Cyber Risk: Integrating Cyber Risk with Financial Risk, White Paper (W215), published by The Open Group, November 2021; refer to www.opengroup.org/library/w215 (the first White Paper of this series)

W221 — Calculating Reserves for Cyber Risk: Vetting Cyber Risk Models, White Paper (W221), published by The Open Group, July 2022; refer to www.opengroup.org/library/w221 (the second White Paper of this series)

Acknowledgements

The Open Group gratefully acknowledges the authors of this document:

- Mike Jerbic
- Robert Mark

The Open Group Security Forum acknowledges the contribution of the following people in the refinement and publication of this document:

- Tony Black, Kyndryl
- Chris Carlson, C T Carlson LLC
- John Feezell, Kyndryl & SRM Working Group Co-Chair
- John Linford, Security Portfolio Forum Director, The Open Group

About the Authors

Dr. Robert (Bob) Mark

Dr. Bob Mark is a Managing Partner at Black Diamond Risk Enterprises. He serves on several boards, has led Treasury/Trading activities, and was a Chief Risk Officer at Tier 1 banks. He is the Founding Executive Director of the MFE Program at UCLA, he has co-authored three books on Risk Management, and holds an Applied Math PhD. Bob was awarded Financial Risk Manager of the Year by GARP, is a co-founder of PRMIA, has published extensively in leading business and finance journals, and is an Individual Contributor in The Open Group Security Forum.

Mike Jerbic

Mike Jerbic is the Founder and Managing Director of Trusted Systems Consulting Group specializing in cyber risk management. He is a retired lecturer in the Economics Department at San Jose State University, and serves as Chair of The Open Group Security Forum. Prior experience includes product development engineering and management at Hewlett Packard, and IT project management consulting. He has authored several articles and book chapters for the American Bar Association as a cyber risk and economics contributor.

About The Open Group

The Open Group is a global consortium that enables the achievement of business objectives through technology standards. Our diverse membership of more than 900 organizations includes customers, systems and solutions suppliers, tool vendors, integrators, academics, and consultants across multiple industries.

The mission of The Open Group is to drive the creation of Boundaryless Information Flow™ achieved by:

- Working with customers to capture, understand, and address current and emerging requirements, establish policies, and share best practices
- Working with suppliers, consortia, and standards bodies to develop consensus and facilitate interoperability, to evolve and integrate specifications and open source technologies
- Offering a comprehensive set of services to enhance the operational efficiency of consortia
- Developing and operating the industry's premier certification service and encouraging procurement of certified products

Further information on The Open Group is available at www.opengroup.org.