Enhancing Risk Calibration Methods

Christian B. Smart, Ph.D., CCEA

Chief Scientist, Galorath Federal

Abstract

Calibration methods such as the Enhanced Scenario-Based Method allow analysts to establish cost risk analyses that are based on objective data. Some methods currently in use rely on the normal and two-parameter lognormal distributions. Empirical data, however, indicates that a three-parameter lognormal is more appropriate for modeling cost risk. We discuss three-parameter lognormal distributions and how to calibrate cost risk using this distribution. We compare the results with traditional calibration to two-parameter normal and lognormal distributions. Calibration methods that have been published to date typically deal with only the system level, but there is a need to calibrate risk at the WBS level. We present a method for calibration at the WBS level that was developed and implemented at Missile Defense Agency that has been very successful.

Introduction

Cost analysts often underestimate cost risk. Early in a project perception of risk is biased low, when over optimism causes risk perception to be lower than is realistic. Also, data necessary to develop an in-depth risk analysis is often not available. Calibration of cost risk analysis to empirical data allows analysts a way to objectively assess risk with limited information.

In the early 2000s multiple authors called into question traditional methods for modeling risk. One of the most prominent of these was Nassim Taleb, whose book the Black Swan, published in 2007, criticized financial models that rely on the normal distribution, including the Black Scholes equation and modern portfolio theory (Taleb 2008). In 2009, Douglas Hubbard published a practical book on risk management that explicitly called for injecting realism in cost risk models by calibrating them to empirical data (Hubbard 2009). In presentations at ISPA-SCEA conferences in 2009 and 2011, Smart showed how to calibrate cost risk to cost growth (Smart 2009, Smart, March 2011, and Smart, June 2011). In 2012, Garvey and others published a calibration method that they call the enhanced Scenario-Based Method (Garvey et al. 2012).

The enhanced Scenario-Based Method is the best-known method for calibrating cost risk analysis to empirical data. This method utilizes a normal distribution or a two-

parameter lognormal distribution to model cost risk. Using a two-parameter lognormal results in representations of risk that indicate that cost can underrun by large amounts with a non-trivial probability. Both Smart and Prince, in presentations at previous ICEAA and ISPA-SCEA conferences have found that cost growth is better represented by a three-parameter lognormal distribution, which accounts for location, as well as scale and shape (Smart 2009, Smart, June 2011 and Prince 2017). The use of a three-parameter lognormal allows the analyst to set a lower bound on cost underruns. We provide examples comparing the three-parameter lognormal for calibration to the two-parameter lognormal and the normal distribution.

Calibration methods that have been published to date typically deal with only the system level, but there is a need to calibrate risk at the WBS level since there is often a dearth of information upon which to calculate a probability distribution. We discuss a method for calibration at the WBS level that was developed and implemented at Missile Defense Agency that has been very successful.

Taxonomy of Risk Analysis Methods

Cost risk methods can be classified as either input-based or output-based. Input-based methods are the more common approach. Input-based methods involve multiple steps. It begins with assessing uncertainty on one or more input variables in a parametric equation. This is convolved with estimating uncertainty through a parametric model to determine risk at the WBS level. This WBS-level risk is then aggregated using a simulation or a method of moments approximation.

Output-based methods involve calculating risk for a riskless point estimate. This approach is not as common but is a very valuable approach. In cases where parametric equations are not available, output-based methods may be the only alternative.

Output-based methods can be based on subjective inputs, empirical data, or a combination of both. The Missile Defense Agency in the 1990s and more recently the Intelligence Community Cost Analysis Improvement Group employed a risk scoring approach that was based on subjective inputs on likelihood and consequence (Gupta 2013). The enhanced Scenario-Based Method (Garvey et al. 2012) and Quick Risk (Smart 2011 and Missile Defense Agency Cost Handbook 2012) are two recent examples of output-based methods that are based on empirical data.

It is also possible to combine the two methods. The Missile Defense Agency developed a method in 2013 that involves subjective expert judgement, and calibrates this judgment to empirical cost growth data (Boone and Crowe, 2013). It was quickly

adopted across the cost directorate and is now the de facto method for assessing risk at the Missile Defense Agency. This will be discussed in detail in a later section.

I use the term "calibration" to refer to output-based cost risk assessment methods that have some basis in empirical data. This is standard nomenclature for measuring financial risk (Hubbard 2009) and better describes the process than other terms that have been used.

The Need for Calibration

Risk is not static over time, but evolves as the project matures. The conventional wisdom is that uncertainty shrinks over time. There is a cone of uncertainty that reduces as the project moves from inception to completion. This makes sense – at the end of the project there is no cost uncertainty, and as you move farther in time from that completion date, the more uncertainty you have about the cost. This is reflected as the cone of uncertainty as illustrated in Figure 1.
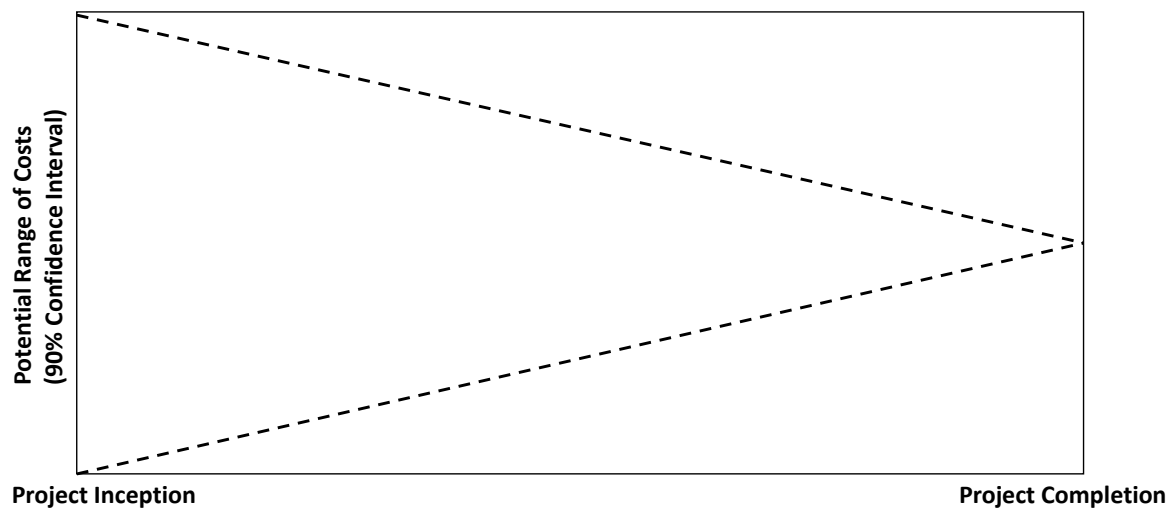

Figure 1. Cone of Uncertainty.

However, this is not typically what happens in practice in terms of how risk is perceived and measured. It is typical for rosy optimism to prevail early in a project's life cycle, with assumptions of a high degree of heritage and few known risks. Later, as the design matures and realism sets in, the perceived risk increases, before narrowing again close to project completion after most of the risks have been realized.

The S-curve is narrower at the beginning of the project when risks are not well understood or their existence is denied. As the project progresses, risks are discovered

and uncovered, resulting in an S-curve that widens. When the project reaches its peak, most of the risks have manifested, resulting in an S-curve being at its widest.

While the risk may be greater at the beginning of the program, the perception is that the risk is low. Many risks are assumed away because of the inherent optimism in the program. It is assumed that a piece of hardware whose design is based on an existing piece of already developed hardware will be easy to build and will encounter no substantial issues in development. The technology readiness of critical parts is assumed to be high. The notion of technological readiness is inherently subjective and subject to over optimism. (Book 2007)

Many of the true risks in the project are not uncovered until the details of the design take shape. As the saying goes, the devil is in the details! For example, early in a project engineers determined that there was a serious technical issue that required a design fix. The identification of these kinds of risks widens the S-curve as they are discovered, leading to an increase in the amount of uncertainty, which widens and flattens the S-curve.

The true way in which risk is measured does not appear to be a cone at all, but more like a diamond. Risk perception starts out narrow, then widens to a peak around critical design review (CDR), then narrows as the project approaches completion.
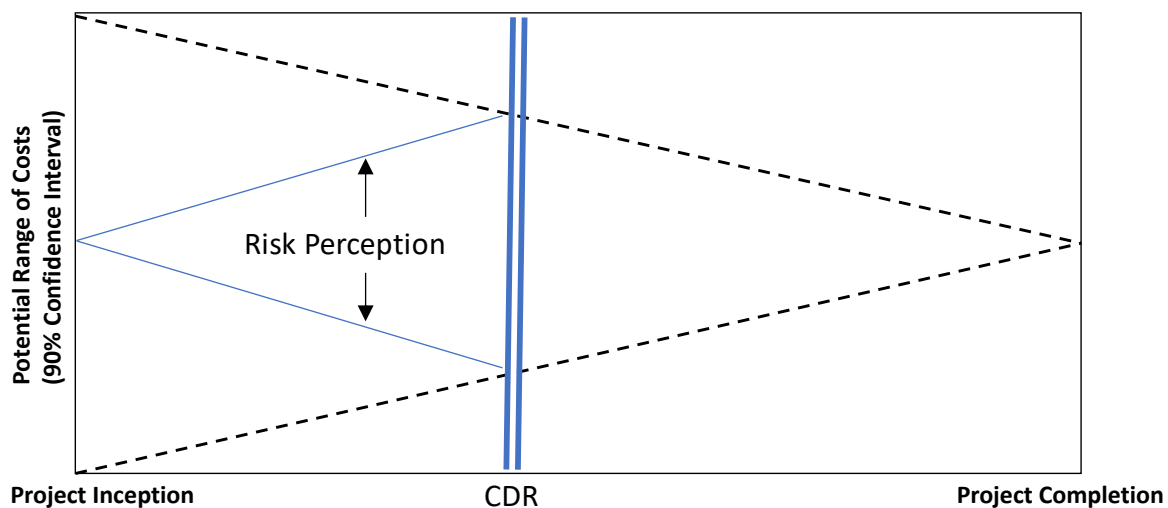

Figure 2. Risk Perception Vs. Reality.

As the physicist Niels Bohr once said "prediction is very difficult, especially about the future." Explaining the past is much easier than predicting the future. However, we confuse our ability to explain the past with our capability of predicting the future. This leads to overconfidence in predicting, which results in an underestimation of risk. Nassim Taleb, author of *The Black Swan*, calls this the "narrative fallacy." (Taleb 2007). Prince and Smart (2018) includes a more in-depth discussion of issues with assessing risk.

In the 1970s, there was a recognition in petroleum engineering that technical people do not understand uncertainty, and that there is a universal tendency to understate risk (Capen 1976). The same holds true today for weapon systems and space projects.

Because of risk blindness and project management pressure to present an optimistic face to upper management, an all-too common situation is that there is a severe disconnect between the cost risk analysis and the final cost. See Figure 3 for an example. Figure 3 displays normalized cost, so the lowest value on the S-curve was assigned a value equal to 1, and the remaining values were normalized based on their value relative to that lowest cost. The Tethered Satellite System (TSS) was a joint project between NASA and the Italian Space Agency (ASI). It consisted of a space tether connected to a 1.6 meter electrically conductive satellite, and was deployed from the Space Shuttle to which the tether was anchored. As the first tethered satellite the project required significant technology development, so it is no surprise that it was inherently risky. And international projects are risky because of the complexity added by coordinating the actions of two separate agencies with diverse cultures.
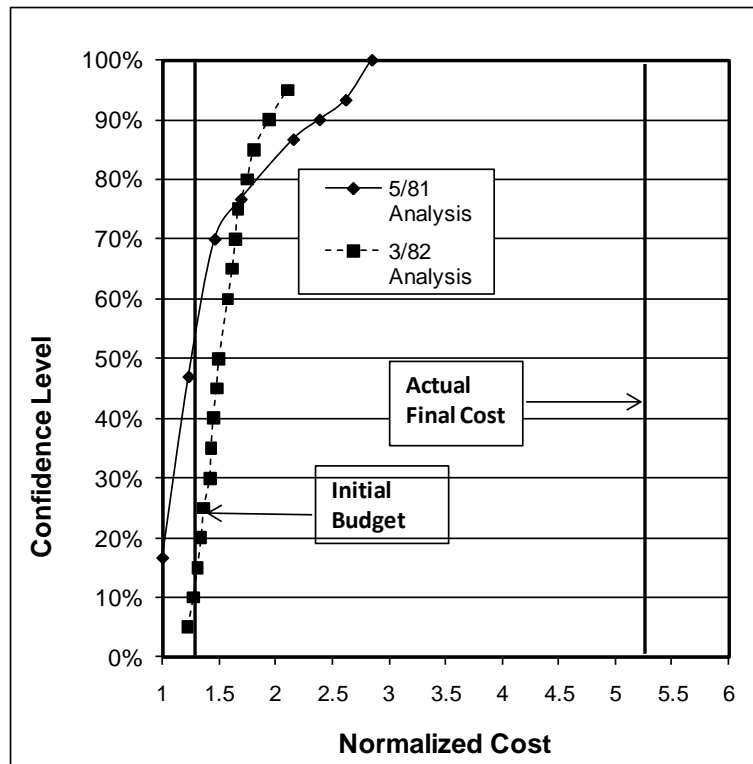
Figure 3. S-curves and Final Actual Comparison for TSS.

Two separate risk analyses were completed at the concept stage for TSS, one in May 1981 and another in March 1982 (cited in Smart, June 2011b). The two S-curves displayed are the results of a Monte Carlo simulation. Note the steepness of these S-curves. It is hard to see much of an "S" in the shape of the second S-curve. It appears to be what has been pejoratively referred to as an "I-beam" rather than an S-curve. Also note that while the initial budget is on both S-curves, the 95th and 100th percentiles from the Monte Carlo simulations are much less than the actual final cost of the project, as the cost growth for this project was extreme, more than 300% from beginning until launch. Note that for the sake of fairness, the analyses were conducted very early in the project's life, as true design and development beyond the concept stage did not begin until 1984.

Like many analyses from that era and more recent ones (Book 2007), the TSS cost risk analyses did not consider correlation, and used limited ranges around the model inputs. Also, model uncertainty, a significant source of risk, was ignored.

Many of the factors necessary for a credible cost risk analysis were included in the risk analysis capability for the NASA/Air Force Cost Model (NAFCOM). With this powerful tool in hand, the author was part of a team that conducted cost risk analyses for a major

project that began in 2006 and proceeded through 2009. Despite accounting for correlation, cost model uncertainty, and variation in the cost drivers with input from project engineers, the Spring 2009 budget was 75% greater than the 50th percentile of the cost estimate developed in 2006, less than three years prior. See Figure 4 for a graph of the S-curve evolution over time compared to the final budget. This was due to a host of factors, some internal to the project, and some external, outside the project's control. Internal factors include early overestimation of heritage (two of the three major elements were supposed to be modifications of existing hardware), and underestimation of the technological challenges. External factors included funding delays, and two major schedule slips.

Note that as posited in a study on how S-curves evolve during a project (Book 2007), the S-curves widened as the project matured, accounting for a greater increase in understanding of the risks involved. Variation in the model inputs widened after receiving additional participation in the risk identification process from project personnel through the implementation of a risk tracking system. But despite incorporating advances in the state of the art in cost risk analysis, the most recent budget was not even on the original S-curve, just as with the TSS curve from 25 year prior. This is partly due to the estimate reflecting project inputs (this was not an independent estimate but rather a project estimate), and a constrained budget
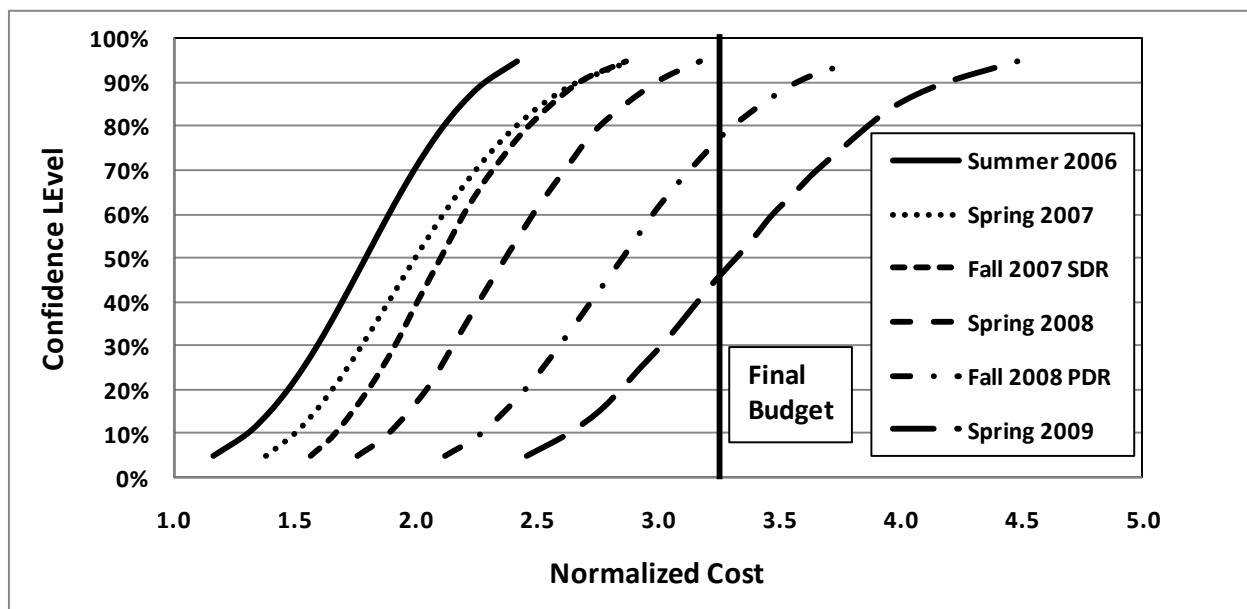

Figure 4. Development S-Curve Evolution for an Actual Project.

environment, which led to non-optimal phasing of cost which led to large increases in cost (Smart 2007). A key point here, however, is that the is that the last S-curve occurred

prior to the Critical Design Review. This is because the project was cancelled by the new Obama Administration – a type of project management black swan! Republican administrations favor human spaceflight while Democratic administrations favor climate science. Still early in the project, we found that as the design took shape, additional risks were uncovered, which led us to widen the S-curve as time progressed. While the actual amount of uncertainty may be greater earlier in the project, the way that we measure the risk it is actually the reverse.

As noted in a study by Steve Book in 2007 relative risk of a project as modeled in numerous cost risk analyses conducted over the previous decade as measured by the ratio of the standard deviation to the mean (which is called the coefficient of variation), ranged anywhere from less than 5% to less than 40%. The cost risk analyses graphed in Figure 4 had coefficients of variation of around 20%, well within the range of typical experience (Book 2007).

Clearly something is still missing from typical cost risk analyses. One way to correct for the early optimism is to calibrate the S-curve to historical cost growth data, especially before CDR.

Calibration is simple. There are a variety of ways to calibrate risk but calibration can be done with as few as two inputs. For example, risk can be calibrated to an initial estimate assuming a percentile for the point estimate and a variation about the point estimate.

Applied to the project illustrated, an S-curve based solely on the initial estimate and calibrated to empirical cost growth data was much more realistic than the more sophisticated and detailed risk analysis that involved uncertainties on the inputs, uncertainty about the CERs, correlation, and Monte Carlo simulation. See Figure 5 for an illustration. The final budget is slightly below the 80th percentile of the empirically calibrated S-curve.
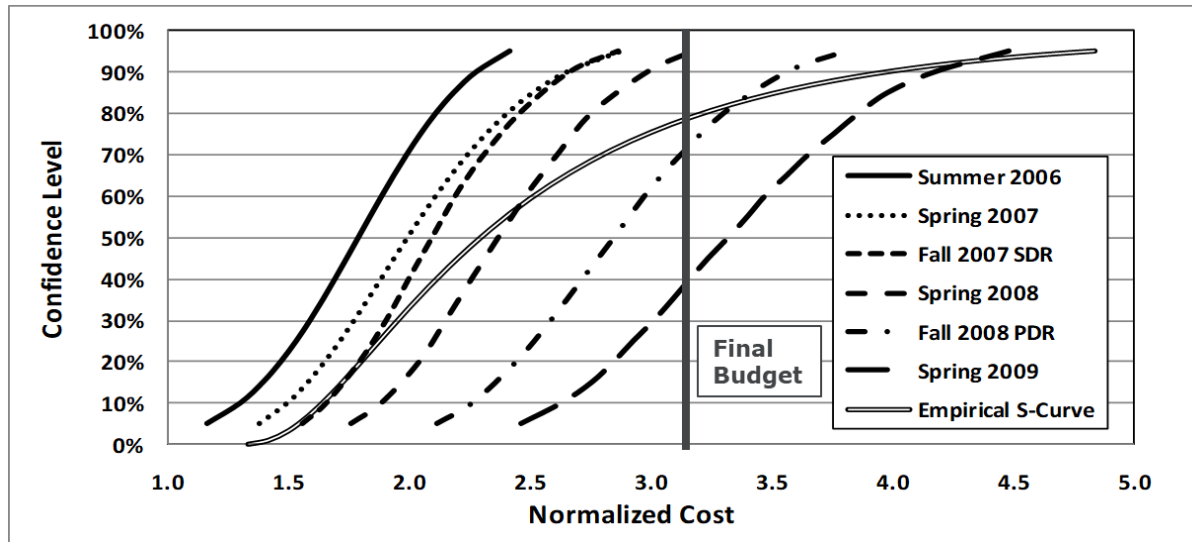
Figure 5. Empirical S-Curve compared to the final budget.

The next sections of the paper go into detail on the history and basics of calibration methods in use, and what we believe is a better way to do calibration.

A History of Calibration Methods in Cost Risk Analysis

The idea of calibration is tied to the key insight that cost growth is cost risk in action. Cost risk is the probability that an estimate will exceed a specified amount, such as $100 million or $150 million. Cost growth and cost risk are thus intrinsically related. While future programs may have risks that are different from those of past programs, and the likelihood and the impact of risks may be different for future programs, historical cost growth provides an excellent means for determining the overall level of risk for cost estimates.

To gain an understanding of how much additional funding will be required to fund a project, in practice it is useful to examine historical cost growth data. Building upon a data set of 112 NASA missions (Smart 2010), a data set of development cost growth was compiled for 289 NASA and DoD programs and projects (Smart 2011). The minimum cost growth was -25.2% for SLWT, a super lightweight version of the Shuttle external tank. The negative number means costs under ran their initial budget by approximately one-quarter of the initial budget (contrary to popular belief, missions occasionally come in under budget.) For the study, 47 missions experienced under runs, which is 16.3% of the missions studied. Only six of the missions hit their budget target spot on. Forty-three of the missions were within 5% of the initial budget, and 70 within 10% (either above or below). The maximum cost growth among the missions studied was 475% for

9

the Comanche helicopter program, which was eventually cancelled before development was completed.

A range from -25% on the low side to over 450% on the high end is a wide range. The average cost growth for all missions was 52.0%, with median growth equal to 29.3%. The difference between the mean and median indicates a high degree of positive skew in the data, with most missions experiencing relatively small amounts of cost growth (half experienced growth less than 30%), with some missions experiencing extreme amounts of cost growth. The data are highly skewed (2.54) with a heavy right tail, as the sample kurtosis is 8.50. Overall, 47 missions had cost growth equal to or in excess of 100%, which means cost at least doubled. While representing only 16.3% of the cost growth data, it has been shown (Smart 2009) that growth of this severity while not extremely common occurs often enough to offset any hoped-for portfolio effect. Indeed many of the issues related to cost growth would be largely ameliorated if project managers could keep cost growth contained within 100%. This would require discipline to contain requirement growth, and realism about the heritage and the technology readiness in the early development stages. See Figure 6 for a graphical summary of these data. The data in Figure 5 include a wide variety of NASA and Department of Defense programs from the 1980s to the 2000s.
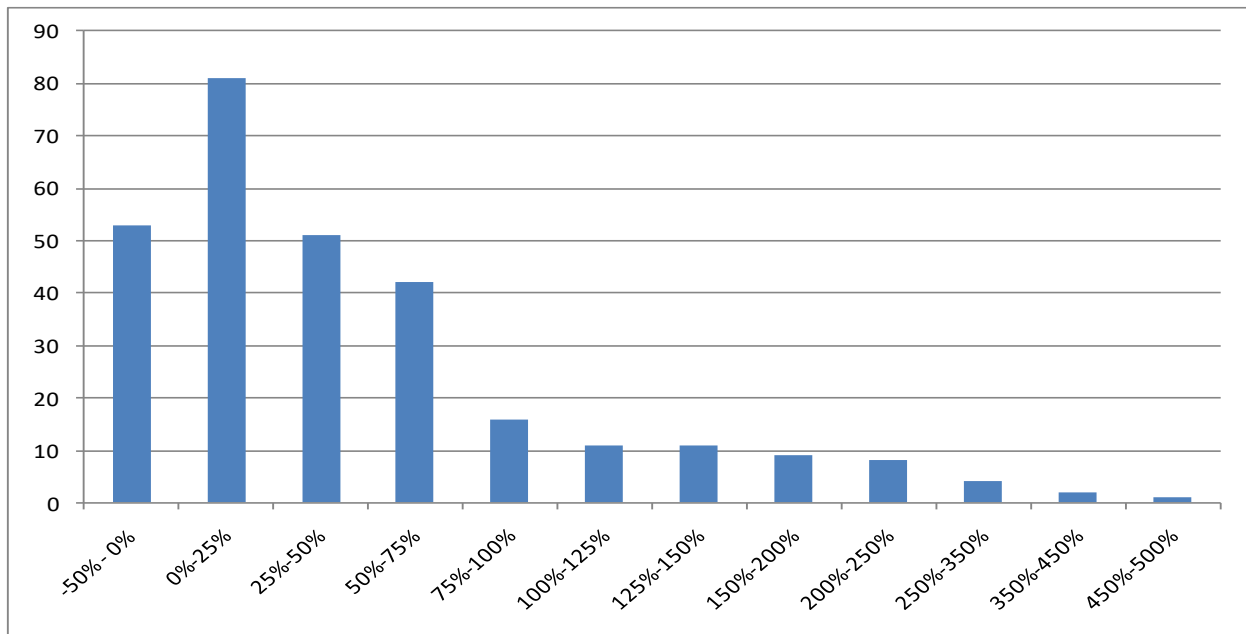

Figure 6. Cost Growth Data for 289 Department of Defense and NASA Programs.

Smart presented a method for calibration of risk to a two-parameter lognormal in March, 2011 (Smart March, 2011), which was later published in the MDA Cost

Estimating and Analysis Handbook (Smart 2012). In June, 2011, Smart presented a method for calibrating risk to a three-parameter lognormal. This paper expands upon that method. In 2012, Garvey and other presented methods for calibrating risk to a normal distribution and a two-parameter lognormal distribution, which they term the enhanced Scenario-Based Method (Garvey et al. 2012).

Two-Parameter Lognormal

A random variable X follows a lognormal distribution if and only if the natural log of X is normally distributed.

The probability density function of a lognormally distributed random variable X is defined as

$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(lnx-\mu)^2}{2\sigma^2}}$$

where μ and σ are mean and standard deviation, respectively, of the log transform of X, i.e., $ln(X)$.

The relationships between the log-space mean and standard deviation and their unit space counterparts, i.e., $E[X]$ and $Var[X]$, are given by

$$E[X] = e^{\mu+\frac{\sigma^2}{2}}$$

$$Var[X] = (e^{\sigma^2} - 1)e^{2\mu+\sigma^2}$$

and

$$\mu = ln\left(\frac{E[X]^2}{\sqrt{E[X]^2 + Var[X]}}\right)$$

$$\sigma = \sqrt{ln\left(1 + \frac{Var[X]}{E[X]^2}\right)}$$

The pth percentile of a lognormal distribution, denoted by α_p, is defined as

$$\alpha_p = e^{\mu+\phi^{-1}(p)\sigma}$$

Where $\phi^{-1}(p)$ denotes the standard normal distribution evaluated at p.

While normally distributed random variable has no lower or upper bound, a lognormally distributed random variable has no upper bound and is bounded below by zero.

See Figure 7 for a notional graph of a lognormal distribution. Note that the lognormal is skewed to the right, unlike the normal which is symmetric and bell-shaped.
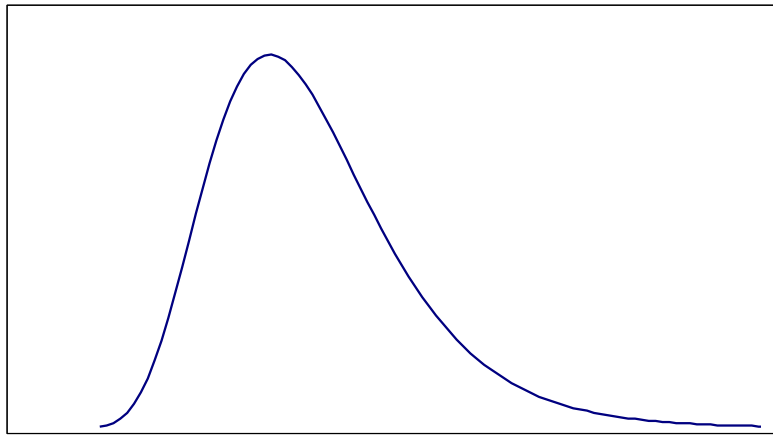

Figure 7. Lognormal Distribution.

Enhanced Scenario-Based Method with the Two-Parameter Lognormal

The enhanced Scenario-Based Method (Garvey et al. 2012) provides equations for modeling risk with both the normal and the lognormal distributions. In this paper we focus only on the lognormal. We do not believe that the normal distribution is suitable for modeling cost risk of government programs. The primary reason is that the normal has thin tails and does not do a good job of modeling the extreme cost growth we often see occur in practice, such as cost growth in excess of 100%. Another reason is because the uncertainty about cost estimates is right-skewed – there are more risks for cost growth than there are opportunities for cost savings, while the normal distribution is symmetric. See papers by Smart (Smart June, 2011) for more information.

The lognormal distribution is a two-parameter distribution. One of the ways to calibrate a lognormal to cost growth history is by specifying a percentile for the point estimate and a coefficient of variation. For example, for a development program at its inception, cost growth studies indicate a CV = 50% and that 80% of program overrun. Thus based on history, we can say we expect that the point estimate will be at the 20th percentile and that the standard deviation will be half the mean. We can use these two facts to calibrate the point estimate to a lognormal.

There are two equations to solve, with two unknowns. The CV determines the log-space standard deviation, σ, since

$$\sigma = \sqrt{ln\left(1 + \frac{Var[X]}{E[X]^2}\right)} = \sqrt{ln(1 + CV^2)}$$

We denote the point estimate by PE. The percentile is defined as

$$PE = \alpha_p = e^{\mu + \phi^{-1}(p)\sigma}$$

Solving this for μ we find

$$\mu = ln(PE) - \phi^{-1}(p)\sigma$$

If you are using Excel, the log-space mean and standard deviations are what you need to as inputs for the LOGNORM.DIST function. If you would rather use the linear space mean and standard deviation, you need to convert μ and σ, which can be done via the equation

$$E[X] = e^{\mu + 0.5\sigma^2}$$

Using the fact that we already know the CV, we can calculate the standard deviation, denoted SD, as

$$SD[X] = CV \cdot E[X]$$

Calibrating the CV with SME Input

When I was at the Missile Defense Agency, we developed a method for calibrating cost risk using subject matter expert (SME) input (Boone and Crowe 2013).

Recognizing that the CV will typically range from 10%-50%, we developed three inputs for calibrating to these values – program definition, experience, and estimating methodology. See Table 1.

Category	Rating	Description
Definition	1	Virtually no definition
	2	Some definition; unclear requirements and exit criteria
	3	Clear definition but missing some requirements and exit criteria
	4	Well-defined with some missing/undefined requirements and exit criteria
	5	Well-defined items, deliverables requirements, and exit criteria; No missing items
Experience	1	Very difficult to estimate; New item/procedure/technology
	2	Difficult to estimate; Item/procedure/technology is 20% similar to previous
	3	Somewhat difficult to estimate; Item/procedure/technology is 50% similar to previous
	4	Easy to estimate; Item/procedure/technology is 70% similar to previous
	5	Very easy to estimate; Item/procedure/technology has been used before; Repeat effort
Estimating Methodology	1	SWAG or heuristic technique
	2	Quote or expert opinion only
	3	Analogy, bottom-up, or parametric with some historical data
	4	Analogy, bottom-up, or parametric with good historical data
	5	Using clear historical data nearly identical to what is being estimated

Table 1. Three Dimensions of Cost Uncertainty.

SME input is used to determine the first two dimensions – definition and experience. The analyst the uses the basis of estimate to determine the estimating methodology factor. Once these three numbers are determined, they are multiplied to obtain the overall rating. The resulting product is then mapped to a CV using the values in Table 2.

Overall Rating	CV	Overall Rating	CV
1	0.50	25	0.31
2	0.48	27	0.30
3	0.47	30	0.29
4	0.45	32	0.28
5	0.43	36	0.27
6	0.42	40	0.26
8	0.40	45	0.24
9	0.39	48	0.23
10	0.38	50	0.22
12	0.37	60	0.21
15	0.36	64	0.20
16	0.35	75	0.18
18	0.34	80	0.15
20	0.33	100	0.13
24	0.32	125	0.10

Table 2. Mapping Uncertainty to CV.

As an example, suppose:

definition = 2 (some definition, unclear requirements);

experience = 3 (50% similar to a previous program); and

estimating methodology = 4 (analogy with good data).

Then the overall rating is equal to 2*3*4 = 24, which results in a CV equal to 32% from Table 2.

Calibrating at the WBS Level

The foregoing discussion has been focused solely on calibrating risk at the system level. We typically do cost estimates at the WBS level, so why not calibrate risk at the WBS level? The cost growth CVs are based on system level cost growth. Unless the common correlation among all elements is equal to *1*, the CVs at the WBS level will be higher.

Given *N* WBS elements, with a common positive correlation coefficient ρ, and common means Mean$_{WBS}$ and standard deviations SD$_{WBS}$, note that the sums of the means and variances at the WBS level are equal to:

$$Mean_{Total} = NMean_{WBS}$$

$$Variance_{Total} = NSD_{WBS}^2 + \rho N(N-1)SD_{WBS}^2 = \left(N + \rho N(N-1)\right)SD_{WBS}^2$$

Thus the total system CV is

$$CV_{Total} = \frac{\sqrt{N + \rho N(N-1)}}{N} CV_{WBS}$$

And the WBS-level CV is

$$CV_{WBS} = \frac{N}{\sqrt{N + \rho N(N-1)}} CV_{Total}$$

As long as ρ is a positive number, this value is approximately

$$\frac{1}{\sqrt{\rho}}$$

For example, if the Total CV is assumed to be 30% and the correlation coefficient is 60%, then the WBS level CV is 1.3*30% = 39%.

Three-Parameter Lognormal

For a constant real number λ, a random variable X follows a three-parameter lognormal distribution if and only if $ln(X - \lambda)$ is normally distributed. X is not bounded above but is bounded below by λ. The value λ is called the location parameter of the distribution.

For a given mean and standard deviation, a three parameter lognormal distribution has the same shape as the two-parameter version, but is shifted by λ. See Figure 8 for an example illustrating that the three-parameter lognormal has the same shape as its two-parameter version, but is just shifted to the left or right.
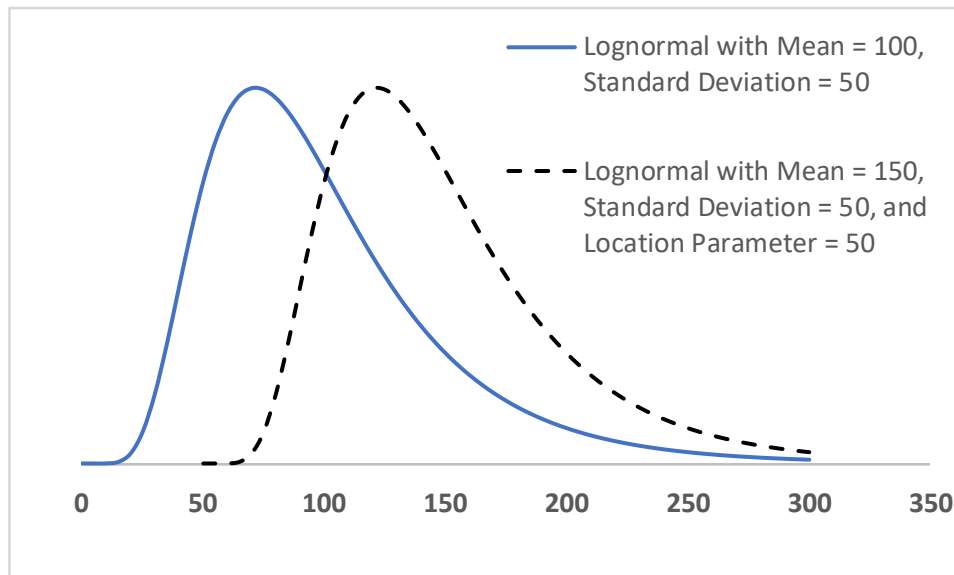

Figure 8. Example of a Three-Parameter Lognormal and a Two-Parameter Lognormal.

The standard deviation in unit space is unchanged by the addition of a location parameter. The coefficient of variation is different, as are the log space mean and standard deviation.

For the three parameter lognormal with location λ, the coefficient of variation is

$$CV[X] = \frac{\sqrt{Var[X]}}{E[X] - \lambda}$$

Also the mean of X has shifted by λ, so we have that

$$E[X] = \lambda + e^{\mu + \frac{\sigma^2}{2}}$$

Thus given $E[X]$ and $Var[X]$, we can find the log-space moments by the following equations:

16

$$\sigma = \sqrt{ln\left(1 + \left(\frac{\sqrt{Var[X]}}{E[X] - \lambda}\right)^2\right)}$$

and

$$\mu = ln(E[X] - \lambda) - \frac{\sigma^2}{2}$$

A three-parameter lognormal is easy to implement in Excel. Once you have determined λ, μ, and σ, you can calculate the value of the CDF at a value $x > \lambda$ as

"=LOGNORM.DIST(x-λ, μ, σ, true)"

Programs for calculating risk such as @Risk include a shift factor capability that can take this into account.

Why a Three-Parameter Lognormal Is Needed

Both Smart (2009, 2011) and Prince (2017) have fit probability distributions to cost growth data, and both have found that a three-parameter lognormal is a better representation of the variation in cost growth than a two-parameter lognormal.

The use of a three-parameter lognormal distribution makes sense. The use of a two-parameter lognormal means that there is a non-trivial probability that cost will decrease by extremely large amounts, such as a 95% decrease, which means that if a project was funded $100 million, there is a possibility that the final cost will be $5 million.

For example, consider a point estimate equal to $100 million that is calibrated to a two-parameter lognormal distribution by assuming that the point estimate is at the *18th* percentile and that the coefficient of variation is equal to 50%, which agrees with my data set. Then the two-parameter lognormal calibration is

$$\sigma = \sqrt{ln(1 + CV^2)} \cong 0.4724$$

and

$$\mu = ln(PE) - \phi^{-1}(0.18)\sigma \cong 5.0376$$

Then the probability of a significant underrun is higher with the two-parameter lognormal than the empirical data. This is because the two-parameter lognormal has a location parameter equal to zero, which means there is some positive probability that underruns will be arbitrarily close to 100%, while the empirical data has a much higher minimum, at -25.2%.

17

In Figure 9, we see that the two-parameter lognormal fit predicts an underrun likelihood probability of less than -30% to be 5%. For all underruns, the two-parameter lognormal overestimates the probability compared to the empirical data. This is easily corrected by using a three-parameter lognormal.
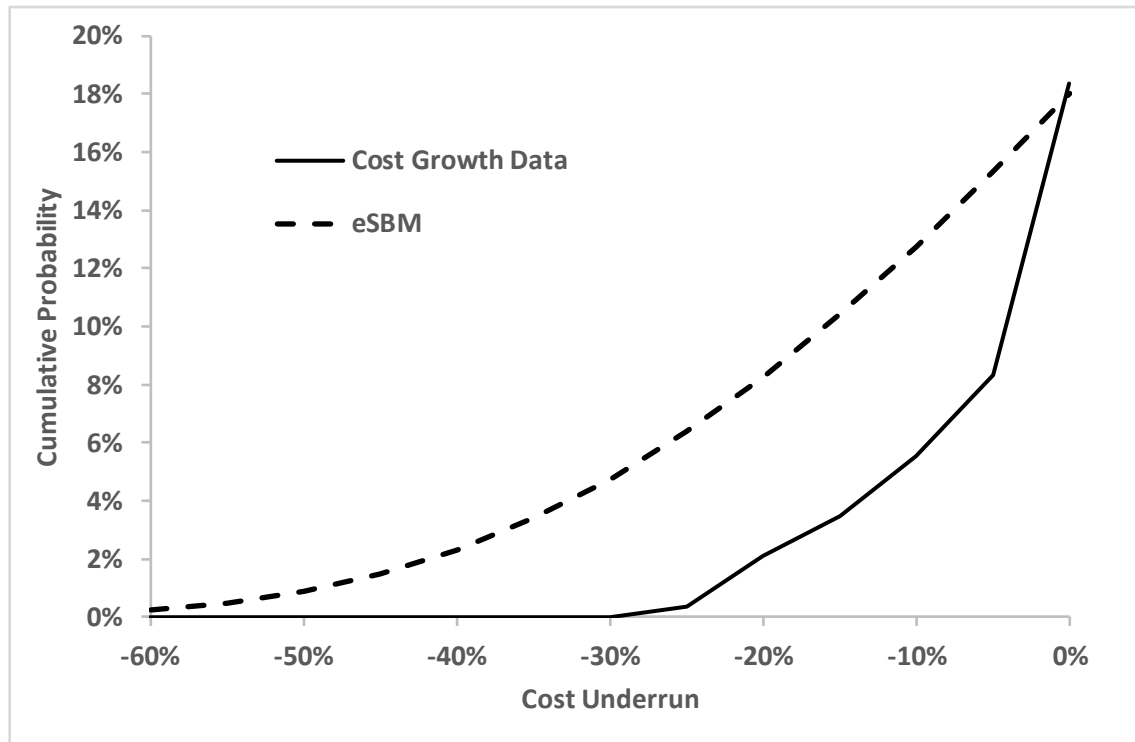

Figure 9. Comparison of Cost Underruns Between Empirical Data and the Enhanced Scenario-Based Model fit to a Lognormal Distribution.

Underruns such as those predicted by the enhanced Scenario-Based Method are unlikely. There is a school of thought that once a program manager receives his funding, he will spend all available budget, and that once a signed contract is in place, the contractor will spend all available funds. The idea is that if not all the funds are spent, the program manager will get a funding cut the next year. Also, many contracts have profit that is tied to their cost, so contractors have no incentive to spend less than the amount on contract. This tendency is referred to as the MAIMS principle – Money Allocated Is Money Spent (Goldberg and Weber 1998). There is no way to model MAIMS directly with a two-parameter lognormal, but we will show how to do it with a three-parameter lognormal.

Also the two-parameter lognormal misses the bulk of the distribution, just like the normal. See Figure 10 for an illustration.
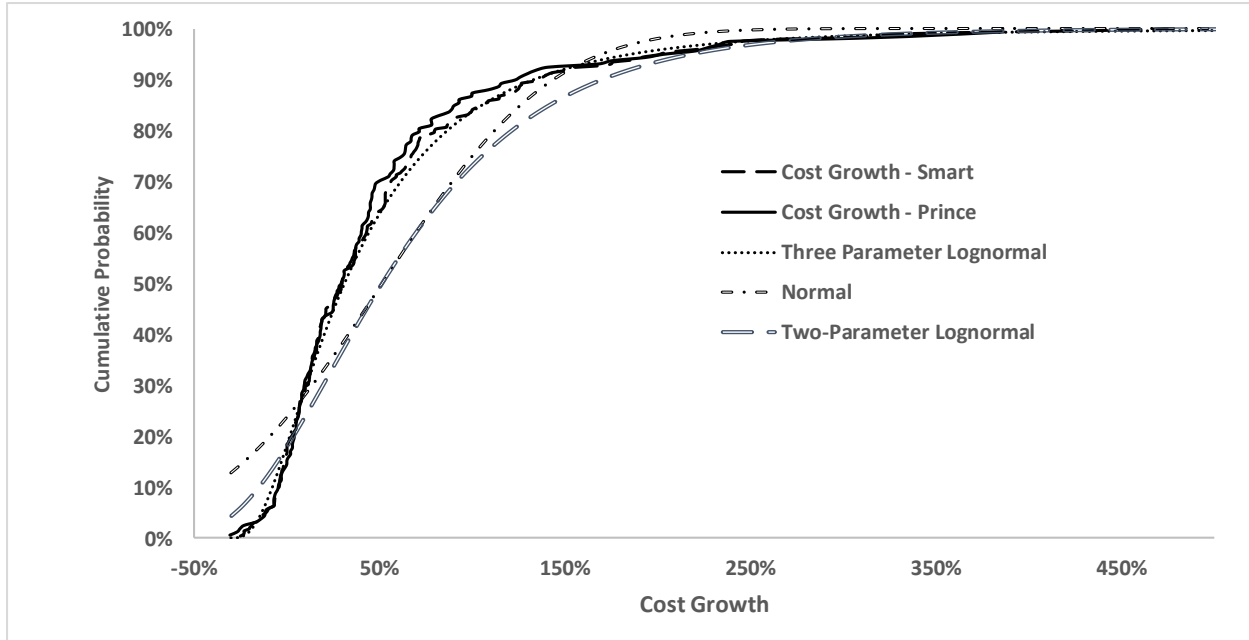
Figure 10. Comparing Cost Growth to Normal, Two-Parameter Lognormal, and Three-Parameter Lognormal Fits.

From Figure 10 it appears that the two-parameter and three-parameter lognormals converge near the right tail, but upon closer examination this does not happen until around the 99th percentile. See Figure 11.

The two-parameter lognormal is not as accurate in depicting the right tail as the three-parameter lognormal, although the two-parameter lognormal tail converges to the three-parameter version at around the 95th percentile.
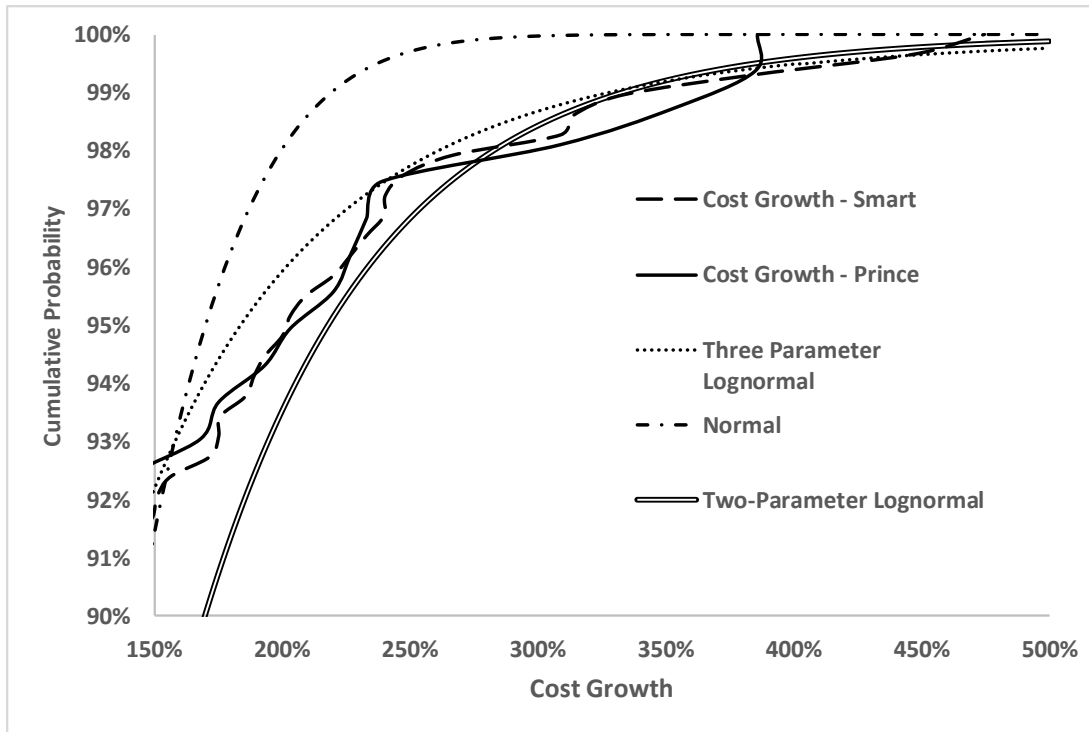
Figure 11. Comparison of the Right Tails of Cost Growth, Normal, and Two- and Three-Parameter Lognormal Distributions.

Calibrating Risk to a Three-Parameter Lognormal Distribution

For the three-parameter lognormal, we need three inputs in addition to the point estimate: a CV that determines the relative risk, the value of the point estimate (percentile, mode, or mean), and the location.

The location parameter is the minimum value for the estimate, a lower threshold. It must be at least zero, and must no larger than the point estimate. A default value based on cost growth data is 30% less than the point estimate, or **0.7*PE**. Other studies indicate a lower threshold such as 50% less than the point estimate (Garvey et al. 2012).

Note that the CV for the three parameter lognormal also involves the location, so it is not the same CV that we have discussed in previous sections and provided guidance for using. However, we can derive the three-parameter CV as needed from the two-parameter CV using empirical cost growth data. The adjustment to the CV will result in different CVs depending upon the location parameter and the assumption of the point estimate (percentile, mode, mean, or location parameter).

Given the percentile and the three-parameter CV we calculate the parameters of the lognormal in the three-parameter case as

$$\sigma = \sqrt{ln\left(1 + \left(\frac{Standard\ Deviation}{Mean - \lambda}\right)^2\right)}$$

$$PE = \lambda + e^{\mu + \phi^{-1}(Percentile)\sigma}$$

Solving for μ we find that

$$\mu = ln(PE - \lambda) - \phi^{-1}(Percentile)\sigma$$

where ϕ^{-1} is the inverse of a standard normal distribution.

As an example, let the point estimate $PE = \$100$, the location parameter $\lambda = \$70$ $million = 0.7*point\ estimate$, the two-parameter $CV = 50\%$, and set the percentile of the PE to the 20th percentile.

Using cost growth studies that indicate that the point estimate is at the 20th percentile and the mean is 1.5 times the point estimate we have that

$$\frac{\sqrt{Var(X)}}{Mean} = 0.5$$

$$\sqrt{Var(X)} = 0.5 * E(X)$$

$$CV = \frac{0.5 * E(X)}{E(X) - \lambda} = 0.5 * \frac{1.5 * PE}{1.5 * PE - 0.7 * PE} = \frac{0.75 * PE}{0.8 * PE}$$

$$\approx 0.9375$$

This is an 87.5% increase from the 50% two-parameter CV.

Then

$$\sigma = \sqrt{ln(1 + 0.94^2)} \approx 0.7957$$

The inverse of the standard normal pdf at the 20th percentile is approximately equal to -0.8416 (this is the same as the z-score from the normal table in an introductory statistics course), so

$$\mu = ln(30) - (-0.8416) \cdot 7957 \approx 4.0709$$

Calculating the linear space mean and standard deviation we find

$$Mean = \lambda + e^{\mu + 0.5\sigma^2} \approx \$150.4\ million$$

$$Standard\ Deviation = (\$150.4 - \$70) * 0.94 \approx \$75.6\ million$$

There are other options than treating the point estimate as a percentile. Three other viable options are to treat the point estimate as the mode, mean, or the minimum value.

Calibrating to the Mode

When calibrating a risk estimate to an analogy, the best choice for the point estimate may not be a percentile, but rather the most likely value, or mode, if the new system is very similar to the historical analogy.

The mode of a two-parameter lognormal distribution is equal to

$$Mode = e^{\mu - \sigma^2}$$

In the three-parameter case, the mode is equal to

$$PE = Mode = \lambda + e^{\mu - \sigma^2}$$

Given the mode and the CV we can solve for the parameters of the lognormal. As before,

$$\sigma = \sqrt{ln\left(1 + \left(\frac{Standard\ Deviation}{Mean - \lambda}\right)^2\right)}$$

Solving for μ in the mode equation yields

$$\mu = ln(Mode - \lambda) + \sigma^2$$

As an example, suppose that the point estimate *PE = Mode = $100 million*, the location parameter *λ = $70 million*, and two-parameter *CV = 50%.*

Cost growth studies indicate that the mode is 5% above the initial cost, the median is 30% higher, and the mean is 50% higher (Smart 2015, Prince 2017).

Thus we assume that the mean is equal to $1.5/1.05 \approx 1.4$ times the point estimate

Thus

$$CV = \frac{0.5 * E(X)}{E(X) - \lambda} = \frac{0.5 * 1.4 * PE}{1.4 * PE - 0.7 * PE} \approx 1.0$$

$$\sigma = \sqrt{ln(1 + 1^2)} \approx 0.8326$$

$$\mu = ln(Mode - \lambda) + \sigma^2 = ln(30) + 0.8326^2 \approx 4.0944$$

$$Mean = \lambda + e^{\mu+0.5\sigma^2} \approx \$154.9 \ million$$

$$Standard \ Deviation = (\$154.9 - \$70) * 1 \approx \$84.9 \ million$$

Calibrating to the Mean

If we believe that the point estimate is equal to the mean, for example, if we have a small number of data points, enough to calculate a mean but not enough to confidently calculate a probability distribution, then the mean may be appropriate for calibration.

As we have discussed, the mean of a three-parameter lognormal distribution is

$$Mean = \lambda + e^{\mu+0.5\sigma^2}$$

Given the mode and the CV we can solve for the parameters of the lognormal. As before,

$$\sigma = \sqrt{ln\left(1 + \left(\frac{Standard \ Deviation}{Mean - \lambda}\right)^2\right)}$$

Solving for μ in the mode equation yields

$$\mu = ln(Mean - \lambda) - 0.5 \cdot \sigma^2$$

As an example, suppose that the point estimate *PE = Mean = $100 million*, the location parameter λ = *$70 million*, and the two-parameter *CV = 50%*.

Then

$$\frac{\sqrt{Var(X)}}{Mean} = 0.5$$

$$\sqrt{Var(X)} = 0.5 * E(X)$$

$$CV = \frac{0.5 * E(X)}{E(X) - \lambda} = 0.5 * \frac{E(X)}{E(X) - 0.7 * E(X)} \approx 1.6667$$

$$\sigma = \sqrt{ln(1 + 1.6667^2)} \approx 1.1529$$

and

$$\mu = ln(Mean - \lambda) - 0.5 \cdot \sigma^2 \approx 2.7366$$

Since the mean is equal to the point estimate, calculating the linear space mean and standard deviation is easy, viz.,

$$Mean = \$100.0 \ million$$

$$\textbf{\textit{Standard Deviation}} = (\$100 - \$70) * \textbf{1.667} \approx \$50.0 \textbf{\textit{ million}}$$

Calibrating with MAIMS

There is a common belief that "money allocated is money spent" (MAIMS)The central idea is that once project managers know how much they have been allocated, they will spend at least that amount, if not more. In the 1990s Lockheed Martin developed a tool to allocate risk based on this principle (Goldberg and Weber 1998). MAIMS does not always hold true, as there are occasionally underruns.

When using the three-parameter lognormal, it is possible to set the point estimate as the minimum value, i.e., *PE = Location=λ.*

Denote the location parameter by *λ.* We need two additional parameters to calibrate the lognormal. Assume a mean value

$$\textbf{\textit{Mean}} = \lambda + e^{\mu + 0.5\sigma^2}$$

The mean value can be based on historical cost growth, for example, at the beginning of development, *Mean = 1.5*PE*.

Assume a coefficient of variation that is the ratio of the standard deviation to the mean

$$CV^* = \frac{\textbf{\textit{Standard Deviation}}}{\textbf{\textit{Mean}}}$$

Then the CV for the three-parameter lognormal is equal to

$$CV = \frac{CV^* \cdot \textbf{\textit{Mean}}}{\textbf{\textit{Mean}} - \lambda}$$

We can calculate the parameters in log space via the following equations:

$$\sigma = \sqrt{ln(1 + CV^2)}$$

$$\mu = ln(\textbf{\textit{Mean}} - \lambda) - 0.5\sigma^2$$

As an example of the application of this approach, let the point estimate *PE = λ = $100* million, with *mean = 1.5*PE = $150 million*, and *CV* = 50%*.

Then the three-parameter CV is equal to

$$CV = \frac{0.5 * 150}{150 - 100} = \frac{75}{50} = 1.5$$

The log-space mean and standard deviation can be calculated as

$$\sigma = \sqrt{ln(1 + 1.5^2)} \approx 1.0857$$

$$\mu = ln(150 - 100) - 0.5 \cdot 0.97^2 \approx 3.3227$$

The linear space mean and standard deviation are

$$Mean = \$150.0 \; million$$

$$Standard \; Deviation = (\$150.0 - \$100) * 1.5 \approx \$75.0 \; million$$

Table 3 compares the mean, standard deviation, and location parameter for the four different three-parameter lognormal calibrations.

	Mean	Mode	20th Percentile	MAIMS
Mean	100.0	154.9	150.4	150.0
Standard Deviation	50.0	84.9	75.6	75.0
Location	70.0	70.0	70.0	100.0

Table 3. Comparison of Four Different Three-Parameter Lognormal Calibrations.

All four calibrations are based on similar assumptions. Note that MAIMS and the 20th percentile calibrations have the most uncertainty, while calibration to the mean is the least conservative. Note also that calibrating to a low percentile, such as the 20th, is similar to MAIMS. Figure 12 contains a graphical comparison of the four different calibrations.
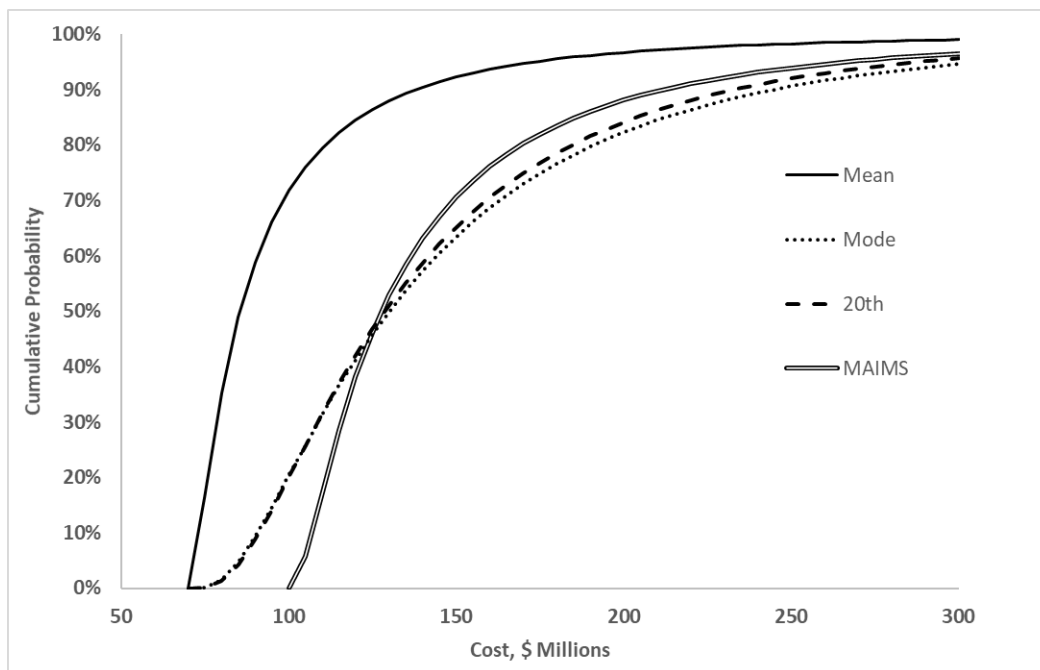

Figure 12. S-curve comparison of four different calibrations to a three-parameter lognormal distribution.

25

Of the four choices for a calibration point, the mean stands out as the most different from the other three. MAIMS is also different in the left tail, but is similar to the mode and 20th percentile calibrations at the 60th percentile and above. The mean calibration results in a much tighter S-curve than the other choices. The takeaway from this chart is that unless you have solid evidence that your point estimate is the mean, such as the output of an unbiased CER, a safer choice is to calibrate to the mode, a low percentile, or use MAIMS.

Summary

Risk perception and risk reality are often out of alignment, especially in the early phases of a project. This is not due to a lack of credible and sophisticated methods for estimating cost risk. Program assumptions influence cost estimates, including the likelihood that cost will increase and the amount that cost will increase. Thus optimistic assumptions and overconfidence early in a program's lifecycle are reflected in the cost risk analysis. Calibration to empirical data is a way to correct for this. Cost growth is cost risk in action – by examining historical cost growth we can develop methods for calibrating cost risk to make it realistic.

Calibration methods to date have focused mostly on two-parameter lognormal and normal distributions (Smart 2011a, Garvey et al. 2012). The normal distribution is not a good choice for modeling risk in most phases of a program, especially for development (Smart 2011b). The two-parameter lognormal also has issues, since there is a lower bound once a contract has been signed. The three-parameter lognormal overcomes the limitation presented by the two-parameter lognormal. Because of this it provides a better fit to historical cost growth data than a two-parameter lognormal distribution. The three-parameter lognormal has been briefly discussed before (Smart 2011b). The present paper expounds on this and provides ways to calibrate to a three-parameter lognormal based on different assumptions for the location of the point estimate, including a percentile, the mode, the mean, and as the minimum bound, as with the MAIMS principle. We recommend calibrating to a percentile, MAIMS, or the mode rather than the mean, unless there is a good reason to believe that your point estimate is at the mean.

Also, calibration methods have focused on the system level. We have presented a way to use SME input to calibrate risk at the WBS level. This method has been in use at the Missile Defense Agency for several years and has proven to be very successful in practice.

References

Book, S.A., 2007, "Cost S-Curves Through Project Phases," NASA Independent Project Assessment Office technical report.

Boone, J.M., and T. Crowe, 2013, "A Scenario-Based Cost Risk Model for THAAD," Missile Defense Agency white paper.

Capen, E.C., 1976, "The Difficulty in Assessing Uncertainty," Journal of Petroleum Technology.

Flynn, B.J., 2011, personal conversation.

Garvey, P.R., B.J. Flynn, P. Braxton, and R. Lee, 2012, "Enhanced Scenario-Based Method for Cost Risk Analysis: Theory, Application, and Implementation," Journal of Cost Analysis and Parametrics,5(2).

Goldberg, M. S., and Weber, C. A., 1998, "Evaluation of the Risk Analysis and Cost Management (RACM) Model." IDA Paper P-3388. Institute for Defense Analysis.

Hubbard, D.W., 2009, The Failure of Risk Management: Why It's Broken and How to Fix It, John Wiley & Sons, Hoboken.

Kleiber, C., and S. Kotz, 2003, Statistical Size Distributions in Economics and Actuarial Sciences, John Wiley & Sons, Hoboken.

Prince, A., 2017, "Being Certain About Uncertainty: Part I," proceedings of the International Cost Estimating and Analysis Association Annual Conference.

Prince, A., and C. Smart 2018, "Being Certain About Uncertainty: Part II," proceedings of the International Cost Estimating and Analysis Association Annual Conference.

Smart, C., 2010 "Here There Be Dragons: Incorporating the Right Tail in Risk Management," Proceedings of the Annual ISPA-SCEA Conference, and 2012, Journal of Cost Analysis and Parametrics.

Smart, C., 2011a, "Quick Turn Risk Analysis," presented to the Greater Alabama Chapter of the Society of Cost Estimating and Analysis.

Smart, C., 2011b, "Covered with Oil: Incorporating Realism in Cost Risk Analysis," Proceedings of the Annual ISPA-SCEA Conference, and 2015, Journal of Cost Analysis and Parametrics.

Smart, C. (ed.), 2012, Missile Defense Agency's Cost Estimating and Analysis Handbook.

Taleb, N.N., 2007, The Black Swan: The Impact of the Highly Improbable, Random House, New York.