COVERED WITH OIL: INCORPORATING REALISM IN COST RISK ANALYSIS

ABSTRACT

“All of those tourists covered with oil” Jimmy Buffett, *Margaritaville*

When Jimmy Buffett sang the words “All of those tourists covered with oil” in his song *Margaritaville* he probably never imagined that this phrase might apply to crude oil instead of suntan lotion. Both the cost and the environmental impact from the 2010 oil spill in the Gulf of Mexico were much worse than anyone had expected or could have predicted. It was, in the words of financial writer Nassim Taleb, a “black swan” – an unexpected event with tremendous consequences. These types of events, like hurricane Katrina in 2005, the giant tsunami in the Indian Ocean in 2004, and the financial crisis that began in 2007 are all examples of events with huge impacts that were hard to foresee. In the arena of government projects, outsized events such as the Challenger and Columbia Space Shuttle disasters cost billions of extra dollars and are not budgeted against. It may be reasonable to not budget for some events that are outside of project management’s control, since doing so will likely lead to excess reserves that go unspent. Unlike natural disasters, project managers have some control over their destiny to the extent that they can meet budget, schedule, and scope by cutting content, and in the cases of extreme overruns, those in authority can cancel projects once they become unmanageable. But budgets for public projects typically include very little risk reserves and do not account for even minimal changes in a project’s design or relatively mild external forces that should be accounted for. In this paper, the author examines historical cost risk analyses and compares them to final actual costs, finding significant differences between the two. Reasons for underrepresentation of risk are discussed, and remedies for this situation are discussed, including the notion of calibration.
ALL COVERED WITH OIL: INCORPORATING REALISM IN COST RISK ANALYSIS

INTRODUCTION

“All of those tourists covered with oil” Jimmy Buffett, *Margaritaville*

When Jimmy Buffett sang the words “All of those tourists covered with oil” in his song *Margaritaville* he probably never imagined that this phrase might apply to crude oil instead of suntan lotion. Both the cost and the environmental impact from the 2010 oil spill in the Gulf of Mexico were much worse than anyone had expected or could have predicted. It was, in the words of financial writer Nassim Taleb, a “black swan” – an unexpected event with tremendous consequences. These types of events, like the recent earthquake and the ensuing tsunami in Japan, hurricane Katrina in 2005, the giant tsunami in the Indian Ocean in 2004, and the financial crisis that began in 2007 are all examples of events with huge impacts that are difficult to foresee. In the arena of government projects, outsized events such as the Challenger and Columbia Space Shuttle disasters often cost billions of extra dollars and are not budgeted against. It may be reasonable to not budget for some events that are outside of project management’s control, since doing so may lead to excess reserves that go unspent. Unlike natural disasters, project managers have some control over their destiny in the sense that they can meet budget, schedule, and scope by cutting content, and in the cases of extreme overruns, those in authority can cancel projects once they become unmanageable. But budgets for public projects typically include very little risk reserves, and do not account for even minimal changes in a project’s design or relatively mild external forces that should be accounted for.

There are numerous types and sources of risk. These include internal factors that are to some extent under a project’s control, such as requirements; external factors, such as strikes and acts of nature; and model uncertainty (models don’t explain the past perfectly). Cost risk analyses typically cover many of these but exclude others. Those typically excluded are extreme events such as the cost consequences of an F-5 hurricane hitting Kennedy Space Center. There are good reasons to leave out some of these. Cost risk analysis is intended to provide decision makers with information to help them successfully manage projects. Inclusion of some extreme events with large impacts will not aid decision makers in budgeting for their projects. Thus exclusion of some risks is advisable. See Figure 1 for a graphical illustration of sources of uncertainty in cost risk analyses (Hunt, 2006).

Estimating uncertainty accounts for the fact that before project completion most facets of a project are only known within a range. There is a projected most likely value, but significant uncertainty exists around project weight, the degree of technology maturation, the amount of heritage that can be leveraged from past projects, the amount of learning that will take place in production, etc. Estimating uncertainty is typically included in cost risk analyses, in the author’s experience with his own cost risk modeling and with those conducted for NASA and
Figure 1. Sources of Uncertainty in Cost Estimates.

Department of Defense agencies. The primary approach is to include ranges around the model inputs by allowing them to vary. This is done by examining historical data for the model inputs or by soliciting expert opinion and engineering judgment on the high and low inputs around the most likely values. These also include such factors as uncertainty surrounding future inflation and test item quantities.

Model uncertainty is the variation in cost attributed to the face that the models that are used to predict cost are not able to explain all the variation in the data. This is due to having limited, finite samples that do not capture every particular scenario, and in some cases captures random effects seen in the past that cannot be systematically captured in a formal model. Also, there may be uncertainty in the model because what is being estimated may not be similar to the data used to develop the cost model. For example, some argue that using robotic satellite cost data to model launch vehicle costs introduces additional uncertainty since there are aspects of launch vehicle systems that are not similar to small robotic satellites. While many of the electronics may be similar at the component level, the structures are significantly different. Also launch vehicles are not produced on a frequent basis, so using older data to estimate a new launch vehicle development program introduces additional uncertainty. These types of uncertainties are often modeled, but not always. Indeed some cost models do not include any model uncertainty in their risk analysis capabilities, hamstringing their users from including this important source of risk.

Uncertainties that are due to hard-to-foresee scenarios include major program re-scopes, and acts of nature. These are typically not included in project risk estimates. This is reasonable, since the purpose of cost risk analyses is to provide project management with useful information for
decision making. These decisions affect those factors within a project manager’s purview, so including only those factors that impact management’s decisions is often ideal for a project estimate. Independent estimates may want to include some of these in order to provide higher-level authorities, such as Congress, with information on what a project may actually cost in practice, since it is likely over the course of a major development that some of these external factors may impact costs across the agency, such as a Shuttle failure impact like the Challenger incident. The Challenger incident in 1986 increased development costs for numerous satellites due to re-designs necessary to find other launch options, and launch vehicle delays, which had a ripple effect that increased costs across NASA into the early 1990s.

While it is good to exclude some categories of uncertainty, we, as a profession, tend to go too far in the opposite direction. Many of these factors, such as model uncertainty, are ignored in developing cost risk estimates. The greater uncertainty inherent when small data sets are used is often ignored as well. And the extent to which requirements will change, the degree of heritage from previous, similar programs that can be relied upon, and the amount of technology development that will need to be conducted are greatly underestimated. Early project plans sometimes have more in common with science fiction than science fact. For example, one satellite project several years ago maintained that it was developing an apogee kick motor for a project that would be a near carbon copy of a previously used apogee kick motor, but the new motor would be twice the size as the “close” analogy. But when looking at the final cost, it was easy to see that a relatively large amount of design cost was required. This inherent optimism of engineers seems to be common, and also reflects the need for project managers to sell their project in the initial stages of a project’s life cycle. This should be taken into consideration when modeling cost risk.

The competitive bidding process has a large impact as well, since contractors know requirements will be changed numerous times during development. They are confident that they can bid low, and then make up for it as change orders are processed and additional money is negotiated for each additional change once cost-plus contracts are signed. As a society, we are risk blind.

“There is a blind spot: when we think of tomorrow we do not frame it in terms of what we thought about yesterday on the day before yesterday. Because of this introspective defect we fail to learn about the difference between our past predictions and the subsequent outcomes. When we think of tomorrow, we just project it as another yesterday.” (Taleb 2007)

Just as the cow who jumped over the moon didn’t think about the risks of re-entering the atmosphere (see Figure 2), so individuals tend to underestimate risk ranges (Air Force 2007). It has been the author’s experience that cost risk analysis tends to be subject to these limitations.

The results of a cost risk analysis are typically presented as an “S-curve,” so named because of the way the graphic typically is displayed. This is a useful format since the confidence levels can easily be read directly from the graph. See Figure 3 for a notional example of an S-curve.
I guess we really didn’t think about the prospect that she would burn up on re-entry.

Figure 2. Risk Not Considered by the “Cow Who Jumped Over the Moon.”

Figure 3. Notional Example of an “S-curve.”

As a result of risk blindness and project 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 4 for an example. Figure 4 displays normalized cost,
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 different cultures.

![Figure 4. S-curves and Final Actual Comparison for TSS.](image)

Two separate risk analyses were completed at the concept stage for TSS, one in May 1981 and another in March 1982 (MSFC 1981, MSFC 1982). 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 actually 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 (MSFC 1992).
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. A number of cost analysts have published
papers stressing the importance of including these in risk analysis (Book 1999, Anderson 2003,
Smart 2005).

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) (Smart 2005). 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 5 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.

![Figure 5. Development S-Curve Evolution for an Actual Project.](image)

Note that as posited in a recent 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 environment, which led to non-optimal phasing of cost which led to large increases in cost (Smart 2007).

As noted in a recent study (Book 2007), the relative risk of a project as modeled in numerous cost risk analyses conducted over the last 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 5 had coefficients of variation of around 20%, well within the range of typical experience.

Clearly something is still missing from typical cost risk analyses. One buzz word that has recently crossed over from finance to cost analysis is the notion of probability distributions with heavy or “fat” tails (Taleb 2007). Fat-tailed distributions are discussed next, along with conclusions from what empirical data suggest what a coefficient of variation would be that is representative of true cost risk.

**FAT-TAILED DISTRIBUTIONS**

Most of the probability distributions that people are familiar with, especially those encountered in basic probability and statistics classes, are those with light or “thin” tails. The normal distribution is a classic example of this. For a normal distribution, 99.7% of the entire distribution is within three standard deviations of the mean. The likelihood of a fluctuation more than six standard deviations away from the mean (either positive or negative) is approximately one in 500 million. This is a perfectly reasonable way to measure random phenomena that fluctuate over relatively limited ranges such as height or age. However, many natural and financial phenomena have fluctuations that are far outside a few standard deviations from the mean. As an example, the 1987 stock market crash was more than 22 standard deviations away from the mean. Thus, the normal distribution is not a good way to model stock market fluctuations. Events whose outcomes extend over extremely wide ranges are modeled with heavy-tailed, or what are sometimes called fat-tailed, distributions. These have received a great deal of press in the financial media in recent years (Taleb 2007). For such distributions the chance of extreme events is much more common than with the normal distribution. Heavy-tailed distributions often have non-finite moments. For example, the first or second moment may not even exist. This means that the standard deviation, and possibly the mean diverge to infinity. A prominent example of a heavy-tailed distribution is the Pareto distribution, which is defined as

\[
    f(x) = \frac{\alpha \theta^\alpha}{(x + \theta)^{\alpha+1}}
\]

for \( x \geq 0 \). The parameter \( \alpha \) determines the degree of heaviness of the tail. When \( \alpha \geq 2 \) the first two moments (mean and standard deviation) are both finite. But when \( 1 \leq \alpha \leq 2 \) only the mean
is finite but the standard deviation is not, and when $0 < \alpha \leq 1$, neither the mean nor the standard deviation is finite. The cumulative distribution function is defined as

$$F(x) = 1 - \left( \frac{\theta}{x + \theta} \right)^{\alpha}$$

See Figure 6 for an example of the cumulative distributions of three Pareto distributions, one fitting each of the three cases.

![Figure 6. Comparison of Three Pareto Distributions.](image)

Note that the three Pareto distributions have drastically different percentiles, although all three bear significant risk. The relative difference between the 70th and 95th percentile for the thinnest-tailed Pareto is a factor of almost four, while for the finite mean, infinite variance version, the difference is greater by an order of magnitude. The relative differences are similar in the infinite mean, infinite variance Pareto, but the 70th percentile is almost an order of magnitude larger for the latter Pareto. See Table 1 for a comparison. The kind of risk exhibited by an infinite mean, infinite variance Pareto is the type of risk seen in box office returns for cinematic films - a mere 5% of total films released earn close to 80% of motion picture profit (De Vany, 2004). Compare how much money *Avatar* has made world-wide compared to the average film. Government projects do not display this degree of risk. Imagine a large $5 billion launch vehicle or missile
development project that suddenly grew to $5 trillion. Such an increase could bankrupt government agencies!

<table>
<thead>
<tr>
<th>Percentile</th>
<th>Finite Mean, Finite Variance</th>
<th>Finite Mean, Infinite Variance</th>
<th>Infinite Mean, Infinite Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>70th</td>
<td>$300</td>
<td>$600</td>
<td>$5,000</td>
</tr>
<tr>
<td>80th</td>
<td>$450</td>
<td>$950</td>
<td>$12,000</td>
</tr>
<tr>
<td>90th</td>
<td>$750</td>
<td>$1,800</td>
<td>$50,000</td>
</tr>
<tr>
<td>95th</td>
<td>$1,150</td>
<td>$3,200</td>
<td>$200,000</td>
</tr>
<tr>
<td>99th</td>
<td>$2,650</td>
<td>$10,500</td>
<td>$5,000,000</td>
</tr>
</tbody>
</table>

Table 1. Comparison of Percentiles for Three Pareto Distributions.

CLASSES OF HEAVY-TAILED DISTRIBUTIONS

The exponential distribution is often used as the dividing line between heavy- and light-tailed distributions. It is considered to be neither light-tailed nor heavy-tailed (Klugman et al. 2008). The probability density function of an exponential distribution is defined as

\[ f(x) = \lambda e^{-\lambda x} \text{ when } x \geq 0. \]

from which the cumulative distribution function (aka “S curve”) can be derived as

\[ F(x) = 1 - e^{-\lambda x}. \]

The tail of the distribution is thus

\[ Pr(X > x) = 1 - F(x) = e^{-\lambda x}. \]

In the case of the exponential distribution, the conditional tail probability has the property that

\[ Pr(X > x + y | X > x) = \frac{e^{-\lambda (x+y)}}{e^{-\lambda x}} = e^{-\lambda y} = Pr(X > y). \]

This is the memory-less property, which means that the conditional probability of a random variable \( X \) being greater than \( x+y \), given that it is greater than \( x \), is equal to the probability that \( X \) is greater than \( y \). As a practical example, suppose that a project’s cost risk is governed by an exponential distribution. This would mean that the risk of the project growing by $100 million would be the same as the probability of the cost growing by another $100 million, given that it had already grown by $100 million. That is, the tail odds do not diminish as cost grows, which should be a hair-raising prospect for a project manager, who would consider such a distribution
to have a heavy-tail indeed! One of the important classes of heavy-tailed distributions is the class of subexponential distributions. Subexponential distributions are so called because their tails are heavier than those of an exponential distribution (Goldie and Kluppelberg 1998).

A simple and intuitive way to compare the heaviness of a distribution’s tail is to compare the limiting behavior of the tail to an exponential distribution, that is,

$$\lim_{x \to \infty} \frac{1 - F(x)}{e^{-\lambda x}} = \frac{1 - e^{-\lambda x}}{e^{-\lambda x}} = \frac{1}{e^{\lambda x}} \int_{0}^{x} \frac{d}{dz} e^{-\lambda z} dz$$

If the limit of the ratio is equal to zero, then the distribution has a thinner tail than an exponential distribution. If the limit diverges to infinity then the distribution has a heavier tail than an exponential. In the case of a normal distribution the limit of the ratio is equal to zero.

Applying L’Hospital’s Rule,

$$\lim_{x \to \infty} \frac{1 - e^{-\lambda x}}{e^{-\lambda x}} = \lim_{x \to \infty} \frac{\frac{d}{dx} e^{-\lambda x}}{\frac{d}{dx} e^{-\lambda x}} = \lim_{x \to \infty} \frac{\lambda e^{-\lambda x}}{\lambda e^{-\lambda x}} = \lim_{x \to \infty} \frac{\exp(\lambda x)}{\exp(\frac{1}{2}(\ln x)^2)}$$

which can easily be seen to be converge to zero.

In this case of a lognormal distribution, assuming without loss of generality that $\mu = 0$ and $\sigma = 1$, then the limit in question is

$$\lim_{x \to \infty} \frac{1 - e^{-\lambda x}}{e^{-\lambda x}} = \lim_{x \to \infty} \frac{\lambda e^{-\lambda x}}{\lambda e^{-\lambda x}} = \lim_{x \to \infty} \frac{\exp(\lambda x)}{\exp(\frac{1}{2}(\ln x)^2)}$$

which is dominated by

$$\exp\left(\frac{1}{2}(\ln x)^2\right)$$
which diverges to infinity. Thus the normal distribution does not belong to the class of subexponential distribution but the lognormal is a subexponential distribution. Thus, normal distributions have thin tails while lognormal distributions have heavy tails. Triangular distributions also have thin tails, while Pareto distributions belong to the class of subexponential distributions.

Subexponentials form a large class of heavy-tailed distributions. Another class of heavy-tailed distributions that is a proper subset of the subexponentials is the class of regularly-varying distributions (Embrechts 2003). Both subexponential and regularly-varying distributions have heavy tails, but regularly-varying distributions have tails that are heavier than some of the distributions in the subexponential class, such as the lognormal. The lognormal distribution is subexponential, but not regularly-varying.

An important subset of regularly-varying distributions is the class of Levy-stable distributions. They are stable in the sense that given two independent, identically distributed random variables, a linear combination of those random variables also has that distributional form. The most well-known example of a stable distribution is the normal distribution. The only other examples are the Cauchy distribution and the class of Levy-stable distributions. Note that an important property of the class of regularly-varying distributions is that they have tail behavior proportional to

\[(1 + x)^{-\alpha}\]

which also describes the tail of a Pareto distribution. Such distributions are thus also called stable Paretian distributions, or \(\alpha\)-stable distributions (Rachev at al. 2007).

There are three commonly-used ways to characterize a probability distribution. The probability density function and the cumulative distribution function (the “S-curve”) are the ones the reader has most likely encountered. Another way to characterize a distribution is through the characteristic function. The characteristic function is a mapping from the set of real numbers into the set of complex numbers, where

\[\Phi(t) = \mathbb{E}e^{itX}\]

is the Fourier transform of the distribution function of the random variable \(X\). Levy-stable distributions, which have four parameters, typically labeled as \(\alpha, \beta, \gamma, \delta\), can only be characterized in a closed-form formula by its characteristic function, and is given by

\[\mathbb{E}(e^{itX}) = \begin{cases} \exp \left[\frac{-\delta|t|}{\alpha} \left[ 1 + i\beta \left( \tan \frac{\pi \alpha}{2} \right) \text{sign} t \left| \frac{\delta}{\alpha} \right|^{1-\alpha} - 1 \right] + i\gamma t \right] & \text{when } \alpha \neq 1 \\ \exp \left[ -\delta |t| \left( 1 + \frac{\beta^2}{\pi^2} \text{sign} t \left( \ln |t| + \ln \delta \right) \right) + i\gamma t \right] & \text{when } \alpha = 1 \end{cases}\]
In the case that $\alpha = 2$ and $\beta = 0$, the Levy-stable distribution is equivalent to the normal distribution (Nolan 1998).

The parameter $\alpha$ determines the tail weight or kurtosis of the distribution and is bound between 0 and 2. The parameter $\beta$ determines the skewness; when $\beta = 0$ the distribution is symmetric. Location is determined by $\gamma$, and scale by $\delta$.

Levy-stable distributions are commonly used to model stock market and other financial market fluctuations, as well as the risk of hurricanes and other natural disasters with potentially extreme consequences.

COST GROWTH DATA AND COST RISK

To gain an understanding of how much additional funding will be required for 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), the author has compiled a data set of development cost growth for 289 NASA and Department of Defense programs and projects (Abata 2004, DAMIR 2010, GAO 1992, 2004, 2009, IDA 2009, Phillips 2004, RAND 2006). The minimum cost growth was -25.2% for SLWT, a super lightweight version of the Shuttle external tank. The negative number means that 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 current 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 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 7 for a graphical summary of these data.
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. Historical cost growth provides an excellent means for determining the overall level of risk for cost estimates. For example, if 95% of past programs have experienced less than 100% growth, we should expect that the ratio of actual cost to the initial estimate should be less than 100% with 95% confidence. Thus cost growth is the impact of cost risk in action. Because of uncertainty in historical data, cost models, program parameters, etc., the term “cost risk” is redundant. Thus characteristics of this cost growth data set determine characteristics seen in a cost risk distribution that is consistent with cost growth.

The cost growth data were fit to a variety of standard probability distributions using Crystal Ball, an Excel add-in. Crystal Ball uses maximum likelihood estimation to fit probability distributions, which works well when a large number of data points are available, as in this case. To assess the fit of the distributions, the Anderson-Darling, Chi-Square, and Kolmogorov-Smirnov (K-S) statistics were calculated for each distribution. The statistics for the top three, as ranked by the Anderson-Darling statistic, are displayed in Table 2.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Anderson-Darling</th>
<th>Chi-Square</th>
<th>Kolmogorov-Smirnov</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lognormal</td>
<td>0.7221</td>
<td>21.6090</td>
<td>0.0471</td>
</tr>
<tr>
<td>Gamma</td>
<td>2.8967</td>
<td>37.6782</td>
<td>0.0736</td>
</tr>
<tr>
<td>Weibull</td>
<td>5.9722</td>
<td>54.7439</td>
<td>0.1330</td>
</tr>
</tbody>
</table>

*Table 2. Comparison of Best-Fitting Distributions for Cost-Growth Data.*

Figure 7. Summary of Cost Growth Data for 289 NASA and Department of Defense Programs.
Each of these tests can be thought of as a measure of deviation from a perfect fit for the data. Thus for all three, a smaller test-statistic value indicates a better fit. These three tests focus on slightly different aspects of a distribution’s fit. Anderson-Darling is focused on the fit at the tails of the distribution, Kolmogorov-Smirnov measures the maximum difference between the actual data and the fitted distribution, and Chi-Square is a sum of squares deviation measure.

Note that the lognormal distribution is the best-fitting distribution according to all three tests. Note that since the two-parameter lognormal is bounded below by zero and the cost growth data includes a significant number of data points with negative values, a three-parameter lognormal distribution is used in this case. The third parameter sets the minimum location for the distribution.

Even though the lognormal has the best rank according to each test, that does not mean we should unequivocally accept the lognormal distribution as a good representative of the underlying data. When it comes to statistics, we can never positively prove a hypothesis such as “the cost growth data fit a lognormal distribution.” We can however disprove hypotheses with data. Thus the best we can hope to do in distribution fitting is to fail to reject a given hypothesis.

For each test, a critical value is determined based on the degrees of freedom of the data. The Anderson-Darling critical value is unique for each distribution. For example, the critical value for a lognormal at the 5% significance level is 0.7520, and the lognormal test-statistic is below that amount at 0.7221, so the lognormal hypothesis cannot be rejected using the Anderson-Darling goodness-of-fit metric. However, the other two distributions have Anderson-Darling goodness-of-fit metrics well above their respective critical values, leading us to reject the gamma and Weibull distributional hypotheses at the 5% significance level. For the Chi-Square test, the critical value given the number of degrees of freedom is higher than the Chi-Square test statistics for the lognormal, so we again fail to reject the lognormal, but we reject the other two distributions.

The Kolmogorov-Smirnov critical value at the 5% significance level is \( \frac{1.36}{\sqrt{N}} = \frac{1.36}{\sqrt{289}} = 0.08 \).

Comparing this to the Kolmogorov-Smirnov statistics shown in Table 2, neither the lognormal nor the gamma can be rejected, but the Weibull distribution is rejected.

Anderson-Darling is sensitive to the fit at the tails (or lack thereof), and the Kolmogorov-Smirnov test is sensitive to maximum departure. The lognormal is the only one that cannot be rejected at the 5% significance level for all three tests. Fitting the tails of the distribution well is important since the right tail is where the risk is located (Smart 2010). Figure 8 shows a graphical comparison of the empirical cost growth distribution and the lognormal fit. The fit at the tails is also good, as is evident from Figure 9.
Figure 8. Empirical Cost Growth Compared to Lognormal Fit.

Figure 9. Comparison of Lognormal Fit with Empirical Data at the Tails.
A lognormal distribution has been found to fit cost growth data well, and because of the intrinsic link between cost growth and cost risk this indicates that cost risk is well represented by a lognormal distribution. However, some in the cost analysis community have posited that cost risk for government projects, along with stock market fluctuations and natural disasters, should follow a Levy-stable distribution (Neatour 2009). To examine this hypothesis, the cost growth data were fit to a Levy-stable distribution, using a freeware program for Windows PCs called STABLE, which is available for download at http://academic2.american.edu/~jpnolan/stable/stable.html.

The STABLE program was written by John Nolan, who is a leading authority on computational issues involving Levy-stable distributions (Nolan 1998). Using the program, a Levy-stable distribution with parameters $\alpha = 1.1402$, $\beta = 1.000$, $\gamma = 0.2210$ and $\delta = 0.1819$ was generated using the maximum-likelihood method. A Kolmogorov-Smirnov goodness of fit test yielded a value equal to 0.0520, below the 5% significance critical value. A Chi-Square test on 17 intervals, approximately equally sized, yield a test value equal to 25.152, less than the critical value at 5% significance. In both cases, the hypothesis that cost growth follows a Levy-stable cannot be rejected. However the goodness-of-fit statistic for both the K-S and Chi-square tests is higher than for the lognormal, indicating that the lognormal is a better fit. This is despite the fact that the lognormal with location parameter has three parameters, while the Levy-stable has four.

Comparing the graphical display of the fit in Figure 10 to that in Figure 9, it is easy to see that the lognormal better fits the empirical data, especially at the tails.

Figure 10. Comparison of Levy-Stable Fit and Empirical Data.
See Figure 11 for a comparison of the fit at the tail between the lognormal and the Levy-stable distributions. It is evident that the Levy-stable tail is heavier than either the lognormal or the empirical data. There are logical, intuitive reasons to expect this to be true. Levy-stable distributions have been found to describe stock market fluctuations, variation in many other financial prices, such as cotton, and in describing natural disasters, such as the financial losses due to hurricanes, and also in box office returns (Taleb 2007, De Vany 2004). These phenomena have in common the fact that they cannot be controlled. The stock market is just too big for one individual, or even a group such as the government, to stop wild swings. And no one can control mother nature. Government projects, however, are subject to limitations. Project managers and other authorities can cut cost through scope changes or cap growth by cancelling projects outright. Indeed many of the worst cases of cost growth represent projects that grew by more than 100% and were subsequently cancelled. Thus a distribution like the lognormal, which represents a state between the wild randomness of the Levy-stable and the mild randomness of the normal distribution (Smart 2008, Smart 2009, Taleb 2007), is a better choice for representing cost risk and cost growth than a Levy-stable distribution.

**Figure 11. Lognormal Better Fit at Tails Than Levy-Stable.**
SOLVING THE S-CURVE CONUNDRUM

The empirical cost growth data can be used to develop project-level S-curves that are in line with reality. Based on the evidence that the lognormal distribution is best for describing cost growth and thus for modeling cost risk, a lognormal distribution is recommended for representing project risk. Note that the parameters of the lognormal distribution fit to the empirical data are location = -31.8%, mean = 51.5%, and standard deviation = 72.1%. Note that a standard two-parameter lognormal distribution with location parameter 0 has the same standard deviation but with a mean equal to 0.515 + 0.318 = 0.833 so the coefficient of variation is actually 86.6%.

In order to calibrate, note that the calibrated S-curve’s log-space standard deviation is completely determined by the coefficient of variation, i.e.,

$$\sigma = \sqrt{\ln(1 + CV^2)} = \sqrt{\ln(1 + 0.866^2)} \approx 0.748$$

In the author’s experience and as shown anecdotally in the examples discussed in this paper, project budgets and cost estimates are similar, since both are based on project inputs, insight, and opinion. Project budgets tend to be on average at the 20th percentile, since roughly 80% of projects experience cost overruns. Project budgets, which are build ups of lower level elements and do not include risk, or correlation between WBS elements, are typically below the mode of a project cost risk analysis, so setting the overall confidence of the calibrate confidence to the 20th percentile of the project cost risk analysis seems to be a reasonable anchoring point.

The 20th percentile of a lognormal distribution, $X_{0.2}$, is equal to

$$X_{0.2} = e^{\mu - 0.8416\sigma}$$

Solving for $\mu$ yields

$$\mu = \ln(X_{0.2}) + 0.8416\sigma$$

Since the lognormal is a three-parameter lognormal, the location parameter means that in order to calculate $\mu$ from a two-parameter lognormal, $X_0$ represents the difference between the 20th percentile and the lower bound, which is 0.318. $X_0$ is the 20th percentile of the project cost risk S-curve multiplied by 0.318, so what is actually used in the calibration is

$$\mu = \ln(0.318X_{0.2}) + 0.8416\sigma$$

S-curve values for a lognormal can be calculated in Excel using the formula

"=LOGNORMDIST(Y, \mu, \sigma)"

where $Y$ varies from 0 to 1. In order to account for the shift, the calibrated S-curve value is calculated as
In order to apply this to the series of marching S-curves shown in Figure 5, the early 2006 analysis occurred earlier than the typical early benchmark from which cost growth is measured for a project, so the Fall 2007 System Design Review analysis, which is one of the early benchmarks in a project’s lifecycle, was used. See Figure 12 for a comparison of the empirically-based S-curve with the project analyses. Notice that the Spring 2009 S-curve has a similar right tail to the empirically-based S-curve, indicating that the latest analysis has incorporated a sufficient degree of risk.

Step-by-step, the calibration routine can be summarized as:

1. Set the project S-curve 20\textsuperscript{th} percentile as the total S-curve 20\textsuperscript{th} percentile – call this $X_\theta$.

2. From the empirical data the log-space standard deviation, $\sigma$, is 0.748.

3. The log-space mean is then calculated as $\mu = \ln(0.318X) + 0.8416\sigma$.

4. The resulting lognormal needs to be shifted to correctly calculate confidence level values. In Excel, this can be accomplished with the formula “=LOGNORMDIST((X-0.682*X_\theta,\mu,\sigma)”

Note that the project risk analyses in Figure 12 can be thought of as conditional S-curves. That is, they are conditional on a specific set of conditions, such as a set schedule, no funding constraints, etc., while the empirical S-curve does not depend upon such a tight range of conditions. By way of comparison, consider a bivariate lognormal cost and schedule S-curve.
like that presented in Garvey (Garvey 2000). In this case cost and schedule each have marginal (unconditional) lognormal distributions. Suppose that the mean of the cost distribution is equal to $100 and standard deviation equals $50. Suppose that the mean of the schedule distribution is equal to 50 months, with standard deviation equal to 10 months with 90% correlation between cost and schedule. When the schedule is fixed, uncertainty is removed from the overall joint distribution, and also from the marginal distribution which is conditional upon this fixed schedule, resulting in a steeper conditional S-curve. For a range of conditional cost S-curves compared to the unconditional cost S-curve, see Figure 13. Notice the strong similarity to Figure 12.

Figure 13. (Unconditional) S-Curve Compared Conditional S-Curves.

Among the sources of uncertainty displayed in Figure 1, project risk analyses tend to incorporate model uncertainty, and to some extent estimating uncertainty, but they include little, if any, of other sources of uncertainty, such as major schedule shifts, funding constraints, and other unexpected events. This may be reasonable for a project manager who needs to restrict his or her attention on those sources of uncertainty subject to control, but wider, more realistic ranges for model inputs should be used in developing cost risk analyses, in order to have an idea of how bad things might become. To do otherwise in the current era of tight budgets and their concomitant schedule slips is to act like an ostrich with its head in the sand, inviting disaster and courting cancellation.

SUMMARY
Project cost risk analyses have tended to significantly underestimate risk. More recent trends such as consideration of correlation and inclusion of model uncertainty in addition to the standard inclusion of variation in cost drivers have helped make more recent analyses more credible. However, these estimates still have tended to under account for risk relative to the amount of cost growth experienced. So simply including correlation and model uncertainty is not a sure means of developing credible risk estimates. Empirical cost growth data provides a means for understanding how much risk projects have experienced, and can provide a means for comparing the amount of risk a project can actually expect to see vice what is predicted from project cost risk models. The empirical cost growth data have been shown to exhibit fat tails, although these tails are not as heavy as found in some other industries, such as stock market prices and financial losses due to hurricanes. This is expected, based on differences in the nature of government projects compared to stock markets and the whims of mother nature, since for government projects, leadership can have a significant influence in ameliorating cost growth, through remedial measures, project re-scoping, or by outright cancellation.

Using empirical cost growth data in calibrating project cost risk analyses to actual projects was demonstrated. This highlighted a strong similarity between project risk analyses and conditional risk analyses, and showed that project risk analyses can be thought of as conditional upon a specific scenario, or relatively limited range of variation. This highlights that project risk often ignores major sources of uncertainty. Some of these are warranted, but decision makers need to realistically assess risk as well. Even if they don’t always have the budget to protect against some risks, they may be able to plan potential alternate courses of action, such as cutting scope. Understanding such risks may also help to enforce discipline, and stress the impact of changing requirements and deviating from project plans.

REFERENCES