Hulett & Associates, LLC
Project Management Consultants

Project Cost Risk Analysis

By
David T. Hulett, Ph.D.[1]

Hulett & Associates, Project Management Consultants

Background

The typical project often overruns its cost estimate. Overruns are common on government and commercial projects, even when changes in the design are taken into account. One reason this happens is because cost estimating traditionally fails to take into account the risk that the work will actually cost more (or less) than provided by even the most competent estimate.

Future estimates are not facts but statements of probability about how things will turn out. Because estimates are probabilistic assessments, costs may actually be higher or lower than estimated even by seasoned professional estimators. The reasons are often causes that are outside the control of the project manager, but may also be endemic to the estimating process, the project strategy or the corporate culture within the project contractor.

A cost risk estimating method is available that provides more accurate estimates of total project cost. The method is based on Monte Carlo simulation. It helps gain better information than traditional methods because it recognizes that project costs are uncertain.

Objectives of a Project Cost Risk Analysis

Cost risk analysis can answer some questions that the traditional estimating method cannot. Included are:

- "What is the most likely cost?" The traditional method assumes that this is the baseline cost computed by summing the estimates of cost for the project elements, but this is not so.

- "How likely is the baseline estimate to be overrun?" Traditional methods do not address this problem.

- "What is the cost risk exposure?" This is also the answer to the question; "How much contingency do we need on this project?"

---

[1] © 2002 by David T. Hulett. All rights reserved. Hulett & Associates, Project Management Consultants are at 12233 Shetland Lane, Los Angeles, CA 90049, (310) 476-7699, www.projectrisk.com and info@projectrisk.com

- "Where is the risk in this project?" This is the same as: "Which cost elements cause the most need for the contingency?" Risk analysis principles can be used to answer this question.

Traditional Project Cost Estimating

To show how a cost risk analysis is done using the Monte Carlo simulation method we will look first at a traditional estimate for a commercial construction project. The traditional estimate has been constructed at a summary level but each of six major components has been estimated. The table below shows that estimate at completion (EAC):

Table 1
Construction Cost Estimate

Cost Component	Estimate
	(000)
Design	1,500
Equipment	5,000
Foundation, Structure	7,000
Piping, Elect., HVAC	1,000
Labor	4,700
Indirects	6,000
Total Project EAC	25,200

In Table 1 the cost estimate is shown as $25,200,000. How likely is it that this project will be built for this amount? In fact, is $25,200,000 even the most likely estimate? To answer these questions we need to examine the uncertainty in these baseline estimates.

Cost Risk Analysis Data Requirements

Nobody should assume that a cost risk could be accomplished without gathering more data. Gathering these data can be a difficult task but the rewards are valuable.

To follow along on the case study, suppose that the risk analyst has chosen well the various project experts who should be interviewed. These experts will probably include the project team and team leaders. They may include experienced project professionals from the company who are not currently assigned to this project. Outside experts are sometimes included, although this is rare except in cases of public projects.

When these people get into a room, the risk analyst asks first about three numbers for each cost component:

- The pessimistic cost estimate. This assumes that everything goes wrong, including failing to achieve the baseline plan and having to do it over again if this is at least 1% likely.

- The optimistic cost estimate. This assumes that everything goes well, and the work will cost less than this estimate only 1% of the time.

- The most likely cost estimate. The temptation is to assume this is the baseline estimate. This is not always so. Often the baseline estimate is a political document, put together to impress the customer. Each estimate is so optimistic that it cannot be achieved without a great deal of luck and maybe a lot of unpaid overtime hours.

The rationale for each of these three is explored and recorded in the notes of the meeting. The rationale is most important because it points to risk mitigation, which is also discussed in the risk interview.

Suppose that this interview has occurred and the following estimates are secured.

Table 2
Low and High Estimates of Construction Costs by Component

Cost Component	Estimate & Most Likely	Low	High
	(000)		
Design	1,500	1,200	2,000
Equipment	5,000	4,500	6,000
Foundation, Structure	7,000	6,200	8,600
Piping, Elect., HVAC	1,000	800	1,600
Labor	4,700	4,500	5,900
Indirects	6,000	5,000	6,700
Total Project EAC	25,200		

The Low and High ranges are not often symmetrical about the estimate. In fact, they often exhibit a greater likelihood for overruns than for underruns. This is in part because there is a natural barrier (zero) to the lowest cost possible and there are many ways the project can run into trouble on the high side.

In Table 2 the Low and High cost columns have not been summed. It is not reasonable to postulate a project in which each and every component of cost underruns or overruns its most likely estimate to the maximum extent possible. If one is over its estimate and another is, too, they will not be at their full-stop excess cost unless they are tightly correlated (more on correlation below).

In the example above, it is assumed that the baseline estimate is the “most likely” cost. In fact, many estimates are not the most likely when the estimators are questioned closely. Sometimes, the risk interview turns up some baseline estimates that should be changed in order to represent the most likely cost. This is one clear benefit of a risk interview, or indeed of any honest and careful scrubbing of the baseline. But, in this example it will be assumed that the baseline was carefully estimated without being “shaded” or biased in any way, and that new information has been recently incorporated in it.

Gathering this information is the most important value that the risk analyst can add to the project. The Monte Carlo method is accepted and simple to describe. The software combines probability distributions correctly. What is sometimes disputed is whether project practitioners can estimate the low, high, most likely and correlation coefficients correctly. Yet, if they cannot, perhaps they are not ready to start the project. If they do not understand the risks and their potential impacts, the baseline plan is in some jeopardy.

Gathering information about cost risk has value itself. Many people indicate during the data gathering interviews that; “This is the first time anyone ever asked me these questions.” “ We should do this on all our projects.” (Maybe this indicates that misery loves company – the risk interview can take several hours and address issues that are unfamiliar at least and uncomfortable or threatening at worst). “Maybe the boss will listen to you, he certainly won’t listen to me.”

Cost Risk Analysis Probability Distribution

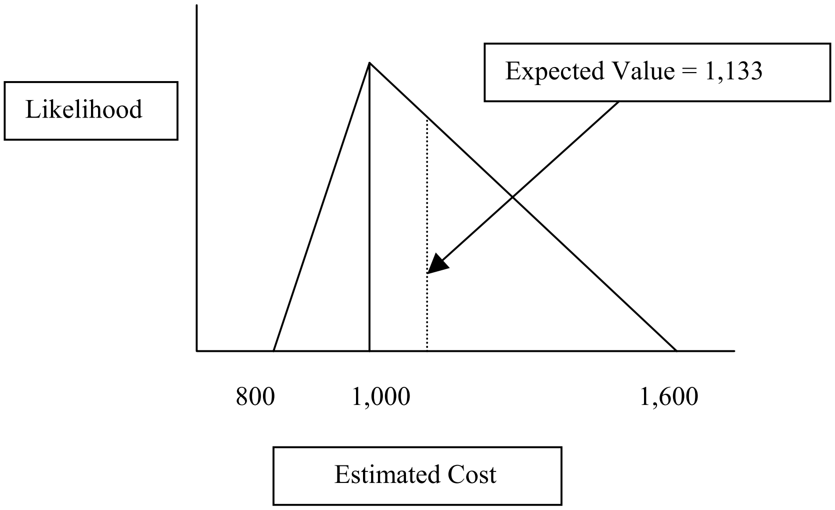
The next item the project risk analyst must discover is the probability distribution shape. Often the triangular distribution is used, but sometimes a different distribution such as the lognormal is assumed. The triangular does not appear in nature but does have several attractive properties.

- Triangular distributions can be completely described by 3-point estimates that are the main purpose of the risk interview. Participants can describe and estimate the low, most likely and high range estimates. It is much more difficult to describe the shape of the curve.

- It is easy to understand and calculate some of the key information about a triangular distribution. For instance, the weighted average or expected cost is found by the equation:

$$\boxed{\text{Expected Value} = (\text{Low} + \text{Most Likely} + \text{High}) / 3}$$

Let us take an example from the case study.



Other probability distributions are available. Some favorites include the Beta distribution and the Log Normal distribution. A custom distribution is also available that the user can use to accommodate special circumstances such as several discrete alternative scenarios. For this example case study the triangular distribution is chosen.

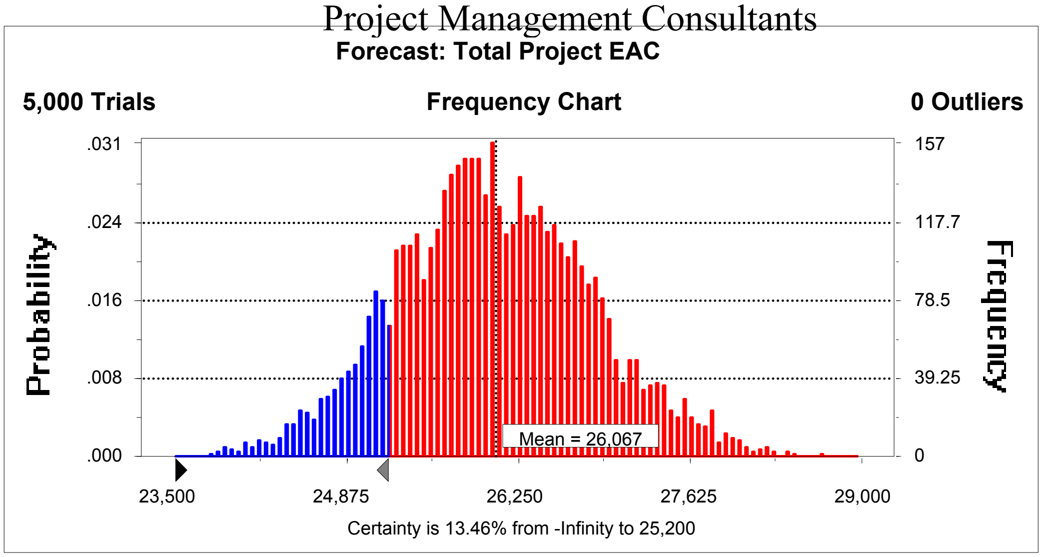
Simulating the Cost Risk Model – Simulation Results

The cost ranges presented in Table 2 above and assuming the triangular distribution will be enough to run a Monte Carlo simulation using Crystal Ball Pro v. 4.0 from Decisioneering. (Other software that simulates Excel is @RISK from Palisade Corporation.) A Monte Carlo simulation "solves" the problem many times. Each solution is called an iteration. For each iteration, the simulation program selects a cost at random from the probability distribution specified by the analyst for each uncertain cost element. The solution in this case is to add them together since this cost estimate is a very straightforward summation model.

One question that needs to be addressed is; "How many iterations is enough?" The answer depends on the use of the information. If the main point of the analysis is to get a more accurate estimate of the most likely cost, fewer iterations are needed. If accuracy in the tails of the result distribution is important, more will be required. Often a stratified sampling method called Latin Hypercube is used to gain more accuracy in the tails for the same number of iterations. In this example 5,000 iterations are used. This is undoubtedly more iterations than needed for most project risk analyses of this simplicity, but it takes very little time on a fast computer.[2]

---

[2] This simulation took about 45 seconds on a 366 MHz computer with a Pentium II and 128 megabytes of RAM.



Three important results can be seen here. Given the risk ranges shown in Table 2:

- First, the EAC (in thousands) of $25,200 is not very likely. The results indicate that the actual cost is about 13.5% likely to be $25,200 or less and 86.5% likely to be overrun. This means that the estimate that was presented to the Board of Directors with such confidence last week is seen as an extremely risky estimate today.

- Second, $25,200 is not even the most likely cost. The most likely cost is about $26,000. This result means that the traditional method of estimating costs by making careful estimates of element costs and then adding them up is just wrong. It is not even the true most likely cost.[3]

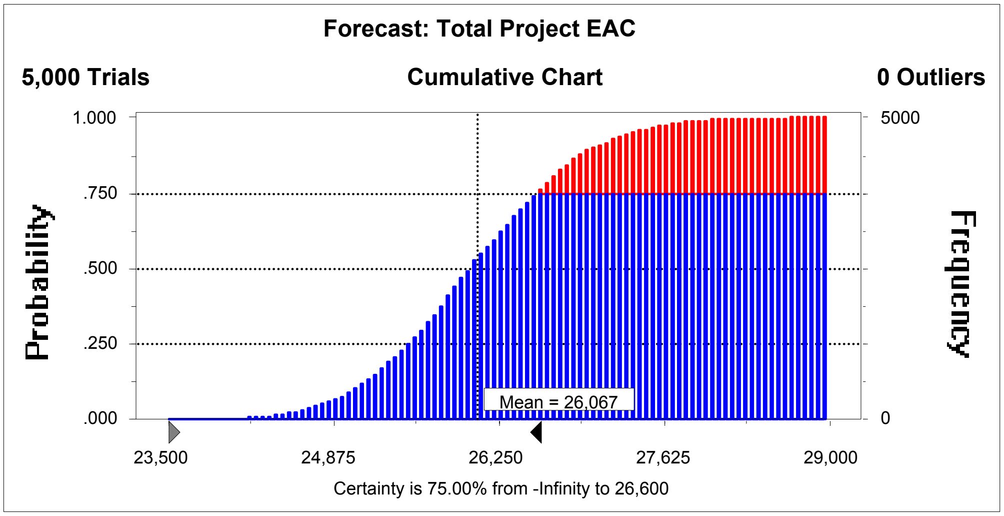
- Third, the mean or weighted average cost is $26,067. This is the sum of the average cost for each cost element, a result we will use below to help identify the high-risk cost elements.

Still, the project manager may ask; "What is my exposure here?" This question expresses the feeling that if the project were 86.5% likely to overrun by a dollar, it does not matter, but if it were even 30% likely to overrun by $20 million, that would be a real problem.

The answer is in the cumulative likelihood distribution. This is just the distribution shown above cumulated from left to right starting at 0 likelihood of cost and cumulating to 100%.

---

[3] See Stephen Book, "Do Not Sum 'Most Likely" Cost Estimates," The Aerospace Corporation, El Segundo, CA, 1994.



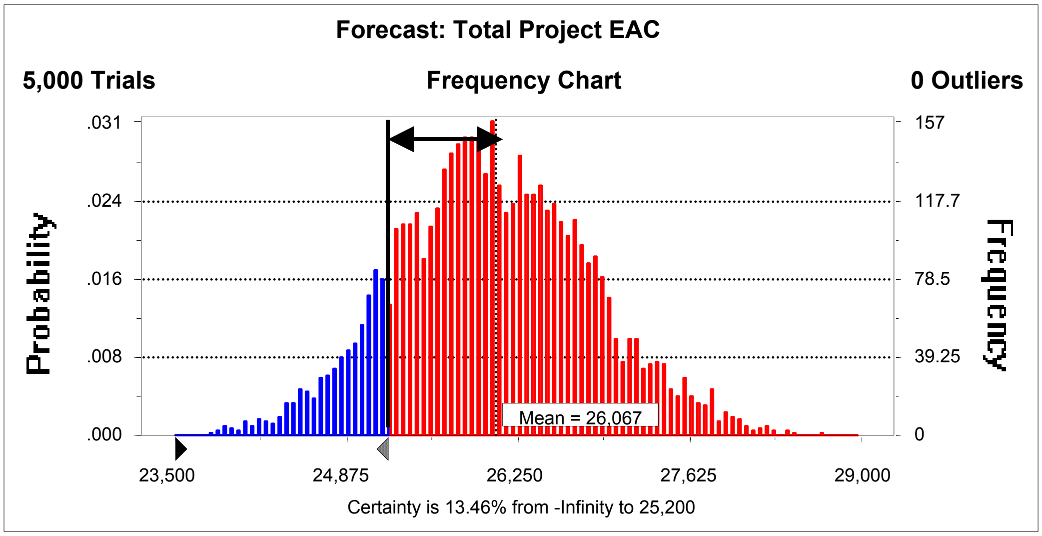
This cumulative chart shows the project manager a vital result. The shading of the amount at 75% cumulative likelihood indicates the first one. Suppose that the company is conservative, and will only bid a project that has at most a 25% likelihood of overrunning. Should they get the bid, this level will be an acceptable risk. The chart shows that the amount bid should be $26,600 (thousands). Another way to look at this result is that a contingency of only 5.6% or $1,400 thousand would be necessary to provide the comfort level of the company. The 86% likelihood of overrun can be corrected easily.

Where is the Risk in the Project?

We can use the principles discovered above to try to identify the high-risk cost elements and rank them by their contribution to risk in the project. Remember that the sum of the mean values of the cost elements is the mean or expected cost for the entire project.[4] This fact is true for simple summation models only. Still, using that fact we can compute the contribution each cost element makes to the need for a contingency *at the mean of the total cost distribution.* The chart below indicates what is being done.

---

[4] This result, that the sum of the means is the mean of the sum, is the basic result of the Method of Moments, an analytical way to approximate the combining of probability distributions. Monte Carlo simulation is a better, more accurate approach to this analysis but reflects this truth.



We need to explain the contingency of (26,067 – 25,200) $867 thousand at the mean. The following table both explains the contingency and ranks the cost elements by their contribution to it. This analysis is especially important to the project manager who may have no intimate knowledge of any of the many cost elements in a complicated project.

Table 3
Construction Cost: Component Contribution to Risk at the Mean

Cost Component	Estimate	Average Cost	Average - EAC
Labor	4,700	5,033	333
Foundation, Structure	7,000	7,267	267
Equipment	5,000	5,167	167
Piping, Elect., HVAC	1,000	1,133	133
Design	1,500	1,567	67
Indirects	6,000	5,900	-100
Total Project EAC	25,200	26,067	867

This table indicates that Labor costs contribute about $333 thousand or about 38% to the contingency needed at the mean. Foundation and structure, and equipment contribute the next most. Indirects are expected to underrun. If the contingency is not at the mean, or if there is correlation (see below) this analysis will not be exact, even in percentage terms, but will be indicative.

Correlation between Project Element Costs

Often project elements' costs are correlated. Correlation is present if the costs of two (or more) cost elements are expected to "move together." That is, two elements' costs will be above or below their average together. Correlation occurs mainly for two reasons:

- The most important is because a common "cost driver" influences both of them. This may be an element of the project, for instance a major technology breakthrough that is assumed to be possible. Several cost elements include activities required to make that technology successful. If the technology proves more difficult or impossible to master, those elements will all cost more. Similarly, if the cost estimate is intentionally "aggressive" or "bare bones" then each cost element may be expected to overrun at the same time. Overruns (and underruns) are reinforced in computing the total project actual cost when positive correlation is strong.

- One element may depend on the other. One example may be subcontracting and the cost of managing that effort. Another may be manufacturing and quality assurance. The support functions are often closely tied to the primary project activity. In this case a direction of causality may be assumed and it might be better to model the dependency rather than to assume a correlation.

The concept of correlation is difficult for many people to grasp. It is defined between pairs of cost elements, although an individual cost element may be correlated with many others. Correlation is defined from –1.0 to +1.0. Negative correlation, when two cost elements move in opposite directions, is not often found in project cost risk analysis. Positive correlation is often assumed, but at what level. Perfect positive correlation is +1.0 and represents two project elements that are always expected to occur in lock step, perhaps reflecting an accounting formula. Correlation above about +0.3 may be useful to include in the modeling, and any correlation above about +0.6 must be considered.

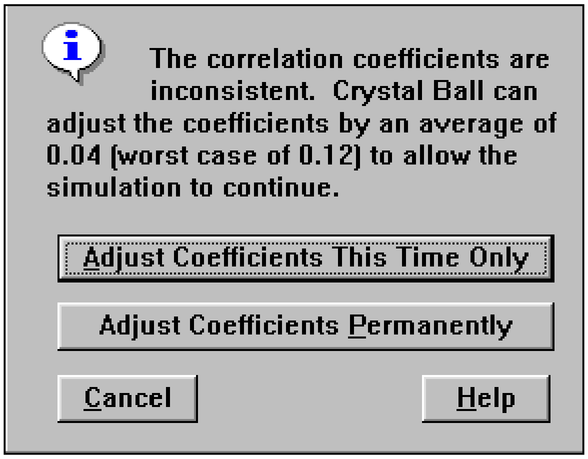
Correlation is implemented in most spreadsheet and schedule applications using the Spearman Rho or rank order approach. This approach bears no known relationship to the standard Pearson product moment correlation. Product moment correlation is what most people do when they specify correlation or what the standard statistics functions (e.g. in Excel) compute when they analyze pairs of numbers. Only one package, Risk+ for Microsoft Project, implements the correct correlation method. It also uses the Lurie-Goldberg approach to adjusting the correlation matrix for non-normal distributions, an advanced topic. It is my observation that the rank order correlation yields results that do not fully reflect correlation as most people think of it.

Unfortunately, there is very little empirical data on correlated cost variables. Usually, the same project participants that estimated the cost ranges will estimate the correlation. Direct estimation of correlation is difficult to do. For instance, suppose that after much discussion about this case study the people estimated the following correlation coefficients:

Table 4
Correlation Coefficients Specified by Judgment

	Design	Equip	Found., Structure	Piping, Elect., HVAC	Labor	Indirects
Design	1.0	0.8	0.7	0.9	0.8	0.9
Equipment		1.0	0.5	0.4	0.6	0.5
Foundation, Structure			1.0	0.4	0.7	0.8
Piping, Elect., HVAC				1.0	0.2	0.4
Labor					1.0	0.6
Indirects						1.0

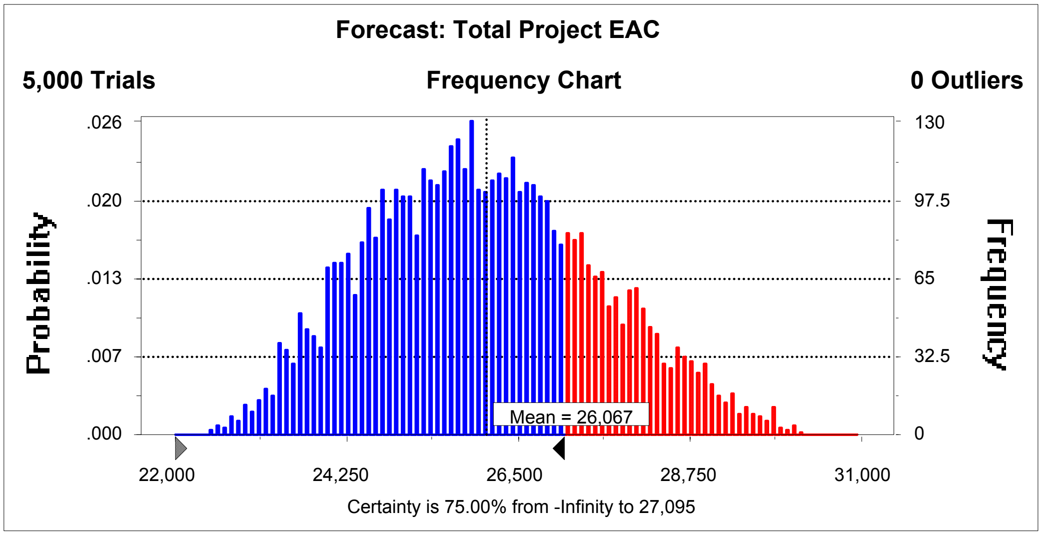
Unfortunately, the project experts have specified a correlation matrix that is logically impossible. Matrix theory implies that a correlation matrix will not have any negative determinants in real life. Yet, with several people sitting around in a room the matrix that results may not be feasible. The software conducts a test, the eigenvalue test, on the correlation matrix to uncover this problem. In this case the group specified many high correlation coefficients and a few low ones. This is illogical, and the program warns the user of the problem.



The correlation coefficients are inconsistent. Crystal Ball can adjust the coefficients by an average of 0.04 (worst case of 0.12) to allow the simulation to continue.

Adjusting the coefficients allows the user to ensure that the correlation matrix is at least not demonstrably impossible. A simple approach to using the correlation algorithm in the program is to adjust the coefficients permanently *after* writing down what they were originally. In this way the analyst will find out after the simulation what Crystal Ball had to do to the coefficients to make them possible. *This is a minimal test and does not ensure that the correlation coefficients are "right" in any sense*. After examining what the program needed to do, the risk analyst still must take responsibility for the coefficients actually used.

The results show how correlation works.



The mean value of the result distribution is the same as before, when correlation was assumed to be absent.[5] This means that the correlation affects the tails of the distribution, which can be seen from the fact that the 75th percentile is now at 27,095 (from 26,600).

The correlation matrix was changed somewhat.

Table 5

Correlation Coefficients after Adjustment

	Design	Equip	Found.Struct.	Piping, Elect., HVAC	Labor	Indirects
Design	1.00	0.75	0.73	0.78	0.72	0.81
Equipment		1.00	0.49	0.42	0.62	0.52
Foundation, Structure			1.00	0.39	0.69	0.79
Piping, Elect., HVAC				1.00	0.24	0.45
Labor					1.00	0.63
Indirects						1.00

Comparing the statistics and the cumulative distributions of the correlated and not correlated runs demonstrates the differences in the simulations.

---

[5] No correlation is when correlation coefficients are set or left at 0.0. This condition that was assumed in the earlier simulation is called "independence."

Table 6
Comparison of Results: Cumulative Distribution

Percentile	No Correlation	With Correlation
0%	23,563	22,495
10%	25,070	24,145
20%	25,403	24,727
30%	25,654	25,171
40%	25,853	25,616
50%	26,045	25,987
60%	26,255	26,414
70%	26,467	26,840
80%	**26,727**	**27,369**
90%	27,085	28,113
100%	29,045	30,258

Clearly, the contingency with correlation assumed needs to be greater at high levels of certainty (e.g. 80%). It is not well understood that the curve extends in the lower direction, too. For instance, the baseline estimate of $25,200 thousand is over 30% likely with correlation vs. 13.5% without. This increase of credibility of the baseline is because positive correlation reinforces the results on both the positive and negative ends. Comparing the statistics of the two runs can show this result.

Table 7
Comparison of Results: Summary Statistics

	No Correlation	Correlation
Trials	5,000	5,000
Mean	**26,067**	**26,067**
Median	26,045	25,987
Mode	---	---
Standard Deviation	**772**	**1,490**
Variance	596,221	2,220,126
Skewness	0.10	0.20
Kurtosis	2.87	2.51
Coeff. of Variability	0.03	0.06
Range Minimum	23,563	22,495
Range Maximum	29,045	30,258
Range Width	**5,483**	**7,763**
Mean Std. Error	10.92	21.07

With correlation the mean is the same but the standard deviation (a measure of dispersion around the mean) and the range width of the result distribution have both increased substantially.

Summary

Traditional cost estimating methods that rely on summing up project elements' cost estimates will not result in even the most likely total project cost. Traditional methods cannot answer the important questions of; (1) how likely are we to overrun? (2) What is our exposure? and (3) Where is the risk in the project? A Monte Carlo simulation can provide answers to these questions.

Data collection is crucial to a powerful risk analysis. Data must be collected on low, most likely and high possible costs and on correlation between elements. Crystal Ball or @RISK would provide the computing power for the Monte Carlo simulations.

The purpose of a cost risk analysis is to assist the project manager by indicating the magnitude of the problem and where risk management efforts should be focused. This paper has indicated how this is accomplished using a simplified construction cost case study.