One Number to Correlate Them All – Efficiently Incorporating Stochastic Dependency in a WBS

Anh Harris, Karen McRitchie, and Christian Smart

Abstract

Correlation is a key consideration in cost and schedule risk analysis, as its exclusion causes significant underestimation of uncertainty. When assigning values in the absence of functional correlation, this can be accomplished by considering every WBS element. However, this can be time-consuming for a detailed estimate. In this presentation, we discuss an alternative method that uses a single value, which offers significant time savings, and discuss its implementation in the SEER® model suite.

Outline

1. Introduction
2. The Importance of Correlation
3. Correlation Modeling Approaches – Modeling One Value Vs. Many
4. What is the Right Value?
5. Implementation of a Single Correlation Value Approach in SEER
6. Conclusion

Introduction

> One value to correlate them all, one value to assign them,
> One value to bring them all, and in the matrix bind them,
> In the Land of Uncertainty, where the shadows lie. (Apologies to J.R.R. Tolkien)

Stochastic dependency in a WBS is a key consideration in cost and schedule risk analysis. One of the most important measures of stochastic dependency is correlation, which is the type of stochastic dependency we will discuss in this paper. A change in the design of one subsystem will impact other subsystems. For example, if the diameter of a launch vehicle increases, additional design work will be required for the structures subsystem, increasing cost for that subsystem. Since the diameter increases, the surface area will also increase, which will require additional thermal protection, which will increase the cost of that subsystem. Also, a change in performance requirements for the entire system may lead to significant re-design for all subsystems, leading to across-the-board cost increases. Schedule slips are common for all types of projects (Smart 2020). Once a project begins development, a schedule slip will result in an increase of costs across multiple or, in many cases, all work breakdown structure elements. These sorts of phenomena are captured in cost risk analysis by applying correlation.

When assigning values in the absence of functional correlation, this can be accomplished by considering every WBS element. However, this can be time consuming for a detailed estimate.

1

In this presentation, we discuss an alternative method that uses a single value, which offers significant time savings, and discuss its implementation in the SEER model suite.

Exclusion of correlation from cost risk analysis results in the significant underestimation of total risk. However, figuring out the correct correlation values between work breakdown structures elements is challenging. For instance, it is difficult to estimate the exact correlation value between the structures and thermal protection subsystems in a cost risk estimate.

The Importance of Correlation

Correlation is often ignored in cost risk analysis. WBS elements can directly influence one another. For example, if the diameter of a missile increases, additional thermal coatings will necessarily be required. Thus an increase in structures cost can directly lead to an increase in thermal control cost. Also, there are underlying common cause factors for cost growth. A budget constraint that leads to an increase in schedule will affect all WBS elements equally.

Consider two random variables, *X* and *Y*. The mean of *X*, *E(X)*, is denoted by μ_x, and similarly, the mean of *Y*, *E(Y)*, is denoted by μ_y. The variance of *X*, *Var(X)*, is denoted by σ_X^2, and similarly, the variance of *Y*, *Var(Y)*, is denoted by σ_Y^2.

The variance of *X* and *Y* are equal to:

$$Var(X) = Cov(X,X) = E(X^2) - [E(X)]^2$$

$$Var(Y) = Cov(Y,Y) = E(Y^2) - [E(Y)]^2$$

Correlation, denoted by the Greek letter ρ ("rho"), is defined by:

$$\rho_{XY} = Corr(X,Y) = \frac{cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{E(XY) - E(X)E(Y)}{\sqrt{Var(X)Var(Y)}} = \frac{E(XY) - \mu_X \mu_Y}{\sigma_X \sigma_Y}$$

The mean, or expected value, determines the central tendency. Variance is a measure of uncertainty about the central tendency. Correlation has a significant impact on the variance. For *n* WBS elements, the mean and the variance of the total cost are defined by:

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} \mu_i$$

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=}^{n} \sigma_i^2 + 2\sum_{j=2}^{n}\sum_{i=1}^{j-1} \rho_{ij}\sigma_i\sigma_j$$

As the variance of total cost is a function of the variances of the individual WBS elements *and* the correlation between them, it is impossible to avoid making a choice about correlation. As in the song "Freewill" by the rock band Rush, "If you choose not to decide you still have made a choice." So the estimator who ignores correlation is making a choice about correlation: the wrong choice, since assuming complete independence will lead to underestimation of total,

aggregate risk. See Figure 1 for a chart of how much the total standard deviation will be underestimated when correlation is assumed to be zero. In Figure 1, n represents the number of WBS elements. For example, for a 100-element WBS, if the actual correlation is 20%, but it is assumed to be zero between all elements, then the total standard deviation will be underestimated by approximately 80%. Note that the amount of underestimation increases with the size of the WBS.
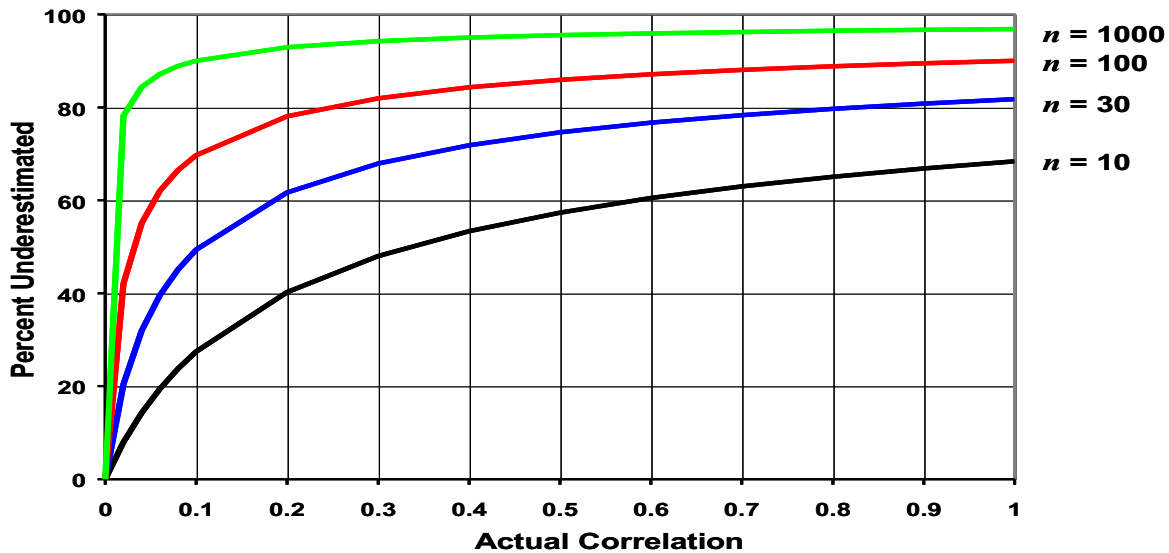

Figure 1 The impact of correlation on risk (Book 1999)

Functional correlation is correlation that is implicit when one cost estimate is a function of another cost estimate. For example, system engineering cost is often modeled as a function of hardware cost. In such a case, when simulation is used to measure and aggregate risk, the variation in hardware cost naturally results in a functional correlation with system engineering cost. In this case, correlation is handled without assigning correlation values and no correlation between hardware cost elements and system engineering needs to be assigned. However, unless structures and thermal control are modeled as functions of a single underlying phenomenon, then correlation between WBS elements needs to be explicitly modeled.

Correlation Modeling Approaches – One Vs. Many

In 1996, Don Mackenzie wrote, "One of the more difficult chores in cost risk analysis is establishing appropriate levels of correlation across the numerous pairs of cost elements in the system cost model." (Mackenzie 1996). More than 25 years later, the cost analysis profession is still struggling with this issue. This paper is an attempt to make progress at addressing this issue.

The difficulty of assigning correlations between WBS elements is one reason for assigning a single value for correlation between all pairs of WBS elements. Another reason is that there are often a large number of correlations to apply. For example, in a 100-element WBS, one has to assign correlation between all distinct pairs, except for the correlation of an item with itself, which is equal to a value of 1.0. Also, correlation is a symmetric measure, so the correlation between elements a and b is equal to the correlation between b and a. Thus, for a 100-element WBS, one has to assign

$$\frac{100^2}{2} - \frac{100}{2} = 4{,}950$$

values. Thus, going through the process of assigning values for each pair can be a time-consuming and tedious process. As an example, consider the notional correlation matrix displayed in Figure 2.

	1	2	3	...	...	n-2	n-1	n
1	1	20%	40%	40%	23%	30%	20%	10%
2		1	30%	50%	5%	50%	40%	10%
3			1	80%	10%	25%	10%	60%
⋮				1	33%	90%	60%	10%
⋮					1	2%	15%	75%
n-2						1	40%	42%
n-1							1	12%
n								1

Figure 2. Notional correlation matrix.

Look at the values in Figure 2. Most analysts will not have the time to fill in all these individual vales nor the knowledge to specify each individual correlation value.

In addition to the time-consuming process of assigning individual values, there is an additional factor to consider. This is that the correlation values cannot be assigned arbitrarily without making the correlation matrix inconsistent. For example, if the correlation between WBS elements 1.1 and 1.2 is 0.9 and the correlation between WBS elements 1.1 and 1.3 is 0.9, then there are constraints on the correlation value between 1.2 and 1.3. Some of the Excel simulation add-ins will correct inconsistent correlation matrices automatically, but it will lead to changes in assigned correlation values that may be significantly different than intended.

Even if two WBS elements have no direct causal connection that could result in correlation, there are underlying factors that often have a significant impact on all WBS elements simultaneously. A prominent example of this is schedule delays. For all types of projects in a variety of different industries, schedule delays occur at least 80% of the time. And when the schedule for a project in development slips, it causes costs to increase for all WBS elements. Thus, it can be argued that it is appropriate to model correlation as if all WBS elements have some amount of positive correlation. As they all have some amount of positive correlation, using a single value makes sense. It is simple, consistent, and is a good approximation for the true underlying correlation structure.

What is the Right Value?

Notice in the graph in Figure 1 there is an apparent knee in the curve around 20%. Above 20% correlation the consequence of assuming less correlation begins to dwindle. This graph is the basis for assuming 20-30% for default correlation for elements between which there is no functional correlation. Book (1999) recommends 20% as a default correlation value because of this. However, the graph in Figure 1 does not tell us how much the total standard deviation is underestimated because correlation is assumed to be 20%, but is actually 60%, for example.

When correlation is assumed to be 10%, but it is actually 20%, the total standard deviation is underestimated. When correlation is assumed to be 40%, but it is actually 20%, the total standard deviation is overestimated. In the simple case for an n-element WBS with all variances and correlations the same, the equation for the total aggregate variance is

$$\sum_{i=1}^{n} \sigma_i^2 + 2\rho \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \sigma_i \sigma_j = n\sigma^2 + 2\rho * n \frac{n-1}{2} \sigma^2 = n\,\sigma^2 (1 + (n-1)\rho)$$

When the correlation ρ is assumed to be ρ_1 but is actually ρ_2 and $\rho_1 < \rho_2$, the percent by which the total standard deviation is underestimated is equal to

$$\frac{\sqrt{n\sigma^2(1 + (n-1)\rho_2)} - \sqrt{n\sigma^2(1 + (n-1)\rho_1)}}{\sqrt{n\sigma^2(1 + (n-1)\rho_2)}}$$

$$= 1 - \frac{\sqrt{(1 + (n-1)\rho_1)}}{\sqrt{(1 + (n-1)\rho_2)}}$$

Similarly, when $\rho_1 > \rho_2$, the percentage by which the total standard deviation is overestimated is equal to

$$\frac{\sqrt{(1 + (n-1)\rho_1)}}{\sqrt{(1 + (n-1)\rho_2)}} - 1$$

For example, for a 100-element WBS, if the correlation is assumed to be 20% but is actually 60%, the total standard deviation is underestimated by 40%. See Figure 3 for a graphical depiction (Book 1999).
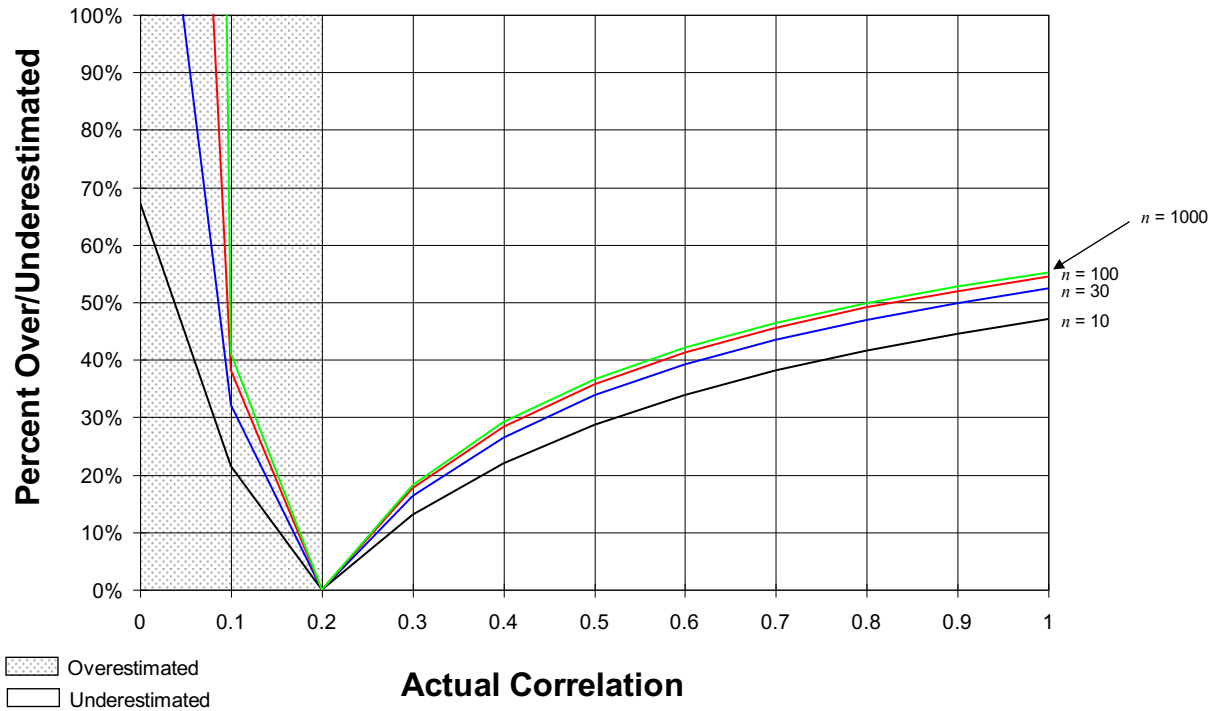

Figure 3. Percent over/underestimated when correlation assumed to be 0.2 instead of ρ (Book 1999)

The knee in the curve approach can still lead to significant underestimation of correlation. We turn to a different approach that seeks to minimize the amount of error in estimation, both over and under.

A more robust approach to assigning correlations would be to use the value that results in the least amount of error. This approach is robust in the sense that without solid evidence to assign a correlation value, it minimizes the amount by which the total standard deviation is underestimated or overestimated due to the correlation assumption. As discussed in Smart (2013), this approach leads to a value of approximately 60%. See Figure 4.
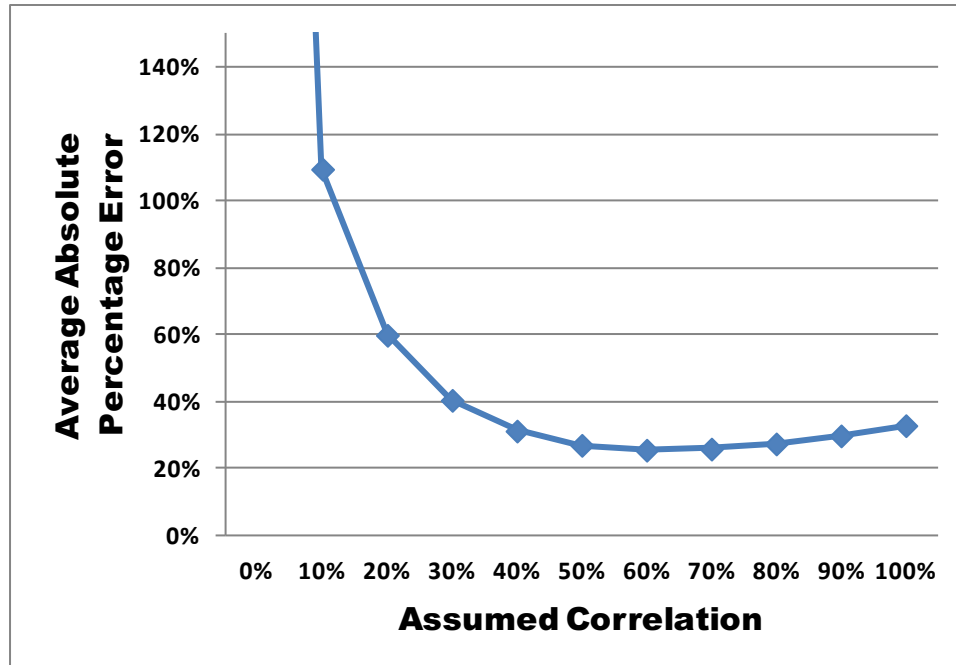
6

Figure 4. Rationale for the 60% value. (Smart 2013, 2020)

The SEER tools are a sophisticated suite of parametric tools that provide a native cost risk capability. Correlation is implemented in the tools. Previously, however, this capability has been limited to either no correlation or 100% correlation across all WBS elements. The ability to choose a single value between 0% and 100% is now implemented in SEER for Hardware (SEER-H™) and SEER for Software (SEER-SEM™), which are discussed in the next section.

Implementation of a Single Correlation Value Approach in SEER

Monte Carlo Basics

The concept of Monte Carlo is to analyze the probability of different outcomes. This is useful for a complex parametric model like SEER which has many inputs and many steps to compute cost.

Determine what is being modeled. In SEER, it is generally cost or hours.

Determine the inputs to the model and how they are distributed. SEER uses a range of inputs (least, likely and most) and applies a modified pert distribution to capture the range. A probability input will determine the value in input range that will be used to compute cost.

Generate a sampling of cost using random inputs for probability. This will involve generating a set of random numbers between 1% - 99% and using those to compute an estimate. Each calculation is an "iteration." You must iterate a sufficient number of times to obtain a statistically significant result.

Compute descriptive statistics from the sampling. Using the sample set generated, SEER computes a mean, standard deviation, and percentiles. This helps to quantify the understanding of the estimate range.

The description above describes the process for a single estimation element. This process can be repeated for multiple elements, with each iteration being added; and the distribution on the total can be evaluated.   Figure 5 illustrates what a simple three element system with a 1000-iteration Monte Carlo might look like. The results on the summed iterations can offer confidence intervals and other statistics that cannot be obtained by simply summing the results of the individual elements.   In this example, the elements are 60% correlated.   Details on how the correlation is considered in the Monte Carlo process is explained in the next section.

Sampling of Development Hours					
Iteration	Element A	Element B	Element C		Sum
1	3,083	4,462	6,415		13,959
2	1,418	2,415	3,615		7,448
3	2,581	3,858	5,597		12,035
4	5,294	7,164	9,683		22,141
5	3,150	4,542	6,522		14,214
:	:	:	:		:
:	:	:	:		:
:	:	:	:		:
:	:	:	:		:
996	1,443	8,438	8,289		18,170
997	920	4,647	14,416		19,983
998	2,572	5,255	14,003		21,830
999	2,636	6,411	8,021		17,069
1000	9,198	4,466	12,407		26,071
Statistics Computed From The Sampling					
Mean	4,549	6,123	8,356		19,028
Median	3,741	5,069	7,300	=/=	17,617
90% Percentile	8,635	10,615	13,692	=/=	29,766

Figure 5 Monte Carlo sampling of 3 elements (1000 iterations, 60% correlation)

Considering Correlation in the Monte Carlo

Historically, SEER offered an all-or-nothing option for correlation. That is, WBS elements could be fully correlated or fully uncorrelated.   More recently, a correlation factor can be specified, the importance of which is described in earlier sections.   The way correlation is considered in SEER is through the random probability draws used as part of the Monte Carlo simulation.

Fully Correlated or 100% Correlation. When elements are fully correlated, the random probabilities used are the same for each element.   If iteration 1 has a random probability of 32%, that will be used for element A, B and C. This means the estimates will move together in their respective ranges and will generally result in a wider range of costs.   One way to think of this is if element A has a relatively bad outcome, elements B and C will also have a relatively bad outcome. Figure 6 shows what the random probability draws would be for each element when there is 100% correlation. Note that each element uses the same set of random probabilities.

Random Probablity Draw - 100% Correlation			
Iteration	Element A	Element B	Element C
1	37.91%	37.91%	37.91%
2	5.01%	5.01%	5.01%
3	26.25%	26.25%	26.25%
4	70.09%	70.09%	70.09%
5	39.46%	39.46%	39.46%
:	:	:	:
:	:	:	:
996	6.30%	6.30%	6.30%
997	49.53%	49.53%	49.53%
998	50.19%	50.19%	50.19%
999	45.10%	45.10%	45.10%
1000	32.87%	32.87%	32.87%

Figure 6. Random probabilities with 100% correlation

Fully Uncorrelated or 0% Correlation. In this case, each element, the random set of probabilities used for each element will be unique. In other words, what happens to element A has no bearing on what might happen to element B or C. The fully correlated case will generally have a tighter range. Figure 7 shows what the random probability draws would be for each element when there is 0% correlation among the 3 elements. Note that each element has a unique set of random probabilities.

Random Probability Draw - 0% Correlation			
Iteration	Element A	Element B	Element C
1	43.91%	34.41%	78.82%
2	27.95%	42.77%	16.43%
3	50.05%	74.93%	92.94%
4	54.50%	78.86%	16.93%
5	93.56%	76.43%	74.72%
:	:	:	:
:	:	:	:
996	94.49%	88.31%	72.76%
997	83.49%	71.75%	65.73%
998	59.40%	8.60%	22.09%
999	94.06%	55.05%	26.20%
1000	92.58%	74.04%	34.38%

Figure 7. Random probabilities with 0% correlation

Partial Correlation. When the correlation falls between 0% and 100%, then there is partial correlation. WBS elements are dependent on each other with a degree "X" % of correlation. When this happens, the random probabilities are the same for all elements for X% of the iterations, and unique for the remaining iterations. For example, with 60% correlation and 100 iterations, SEER will apply 60 iterations of identical randomly drawn probabilities to all WBS elements and 40 iterations of different randomly drawn probabilities to each element. Figure 8 shows what the random probability draws would be for each element when there is 60% correlation among the 3 elements. For the first 60% of iterations, the elements use the same probability. For the remaining iterations, each element has a unique set of probabilities.

Random Probability Draw - 60% Correlation				
	Iteration	Element A	Element B	Element C
	1	37.91%	37.91%	37.91%
	2	5.01%	5.01%	5.01%
	3	26.25%	26.25%	26.25%
60% of iterations	4	70.09%	70.09%	70.09%
	5	39.46%	39.46%	39.46%
	:	:	:	:
	:	:	:	:
	599	51.55%	51.55%	51.55%
	600	51.81%	51.81%	51.81%
	601	89.49%	37.26%	36.50%
	602	94.35%	77.86%	30.99%
	:	:	:	:
	:	:	:	:
	996	5.30%	78.44%	59.43%
	997	1.14%	41.49%	90.07%
	998	26.05%	51.81%	89.05%
	999	27.53%	63.79%	57.08%
	1000	91.11%	38.00%	84.04%

Figure 8. Random probabilities with 60% correlation

Figure 9 shows the range of estimates for different correlation levels. As expected, the fully correlated case has the greatest range while the fully uncorrelated case has the smallest range.
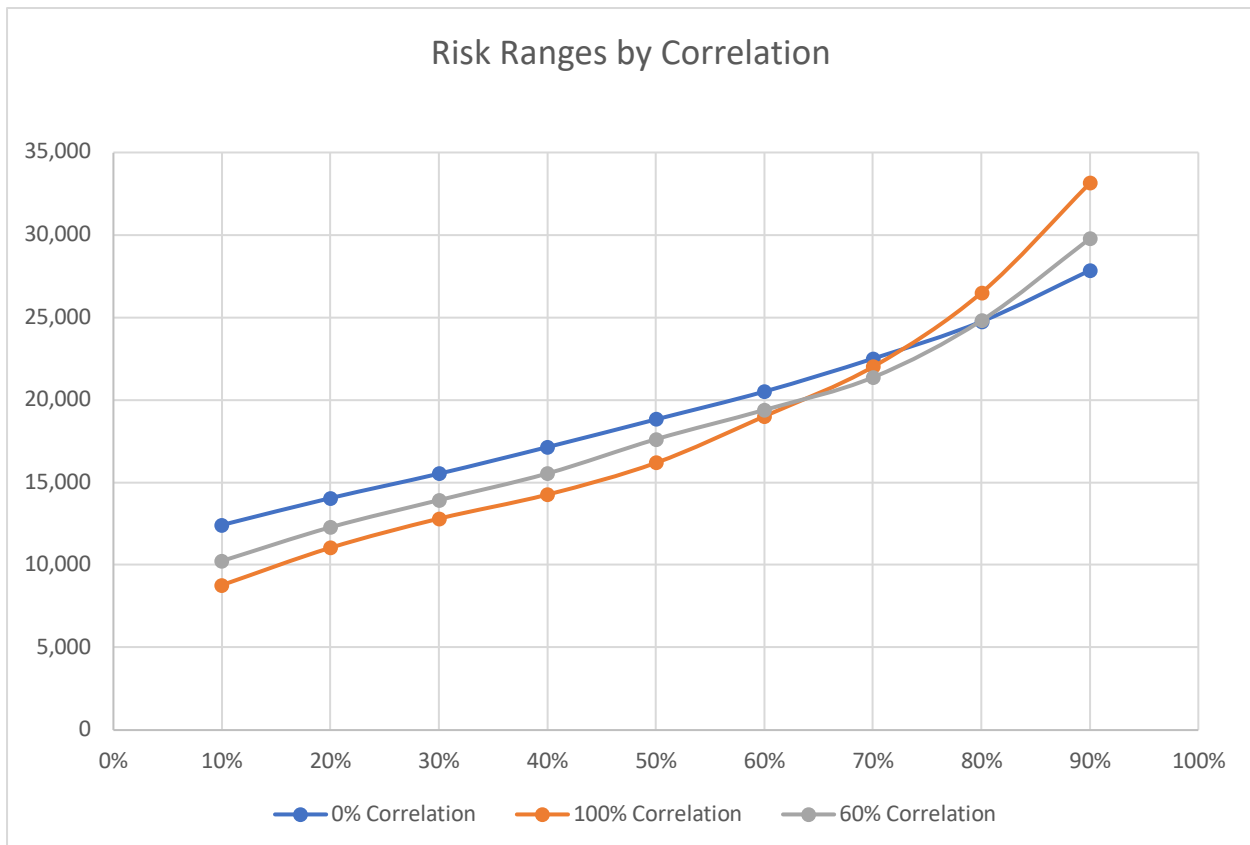

Figure 9. Monte Carlo results for different correlations

Randomness and Repeatability.   By its nature, each Monte Carlo run will result in a slightly different result. This is because each run will use a different random number sequence to drive

10

the result. These differences become very small as you increase the iterations, however they will not go away entirely. Even if the principles are sound in doing so, having slightly different results without changing any assumptions (correlation, iterations) is not ideal for a costing tool used to help make decisions. Because of this, SEER has seeded the random numbers used to ensure repeatable results.

SEER applies the random seed method to initialize randomly drawn numbers in sampling. This ensures that the risk results are repeatable for a given WBS element within an estimate. Internally, each WBS element is encoded with a Universally Unique Identifier or UUID which is essentially a 36 digit alphanumeric string designed to be unique. This string is hashed to create a random number to be used as a seed to initialize random draws for that element. Even if you duplicate a WBS it will be assigned a new UUID and the Monte Carlo analysis will produce different results. The difference in results will grow smaller as the number of Monte Carlo iterations increases.

Figure 10 illustrates how two identical elements could have slightly different Monte Carlo results because they use a different set of random probabilities. These differences become smaller as you increase the number of Monte Carlo iterations.
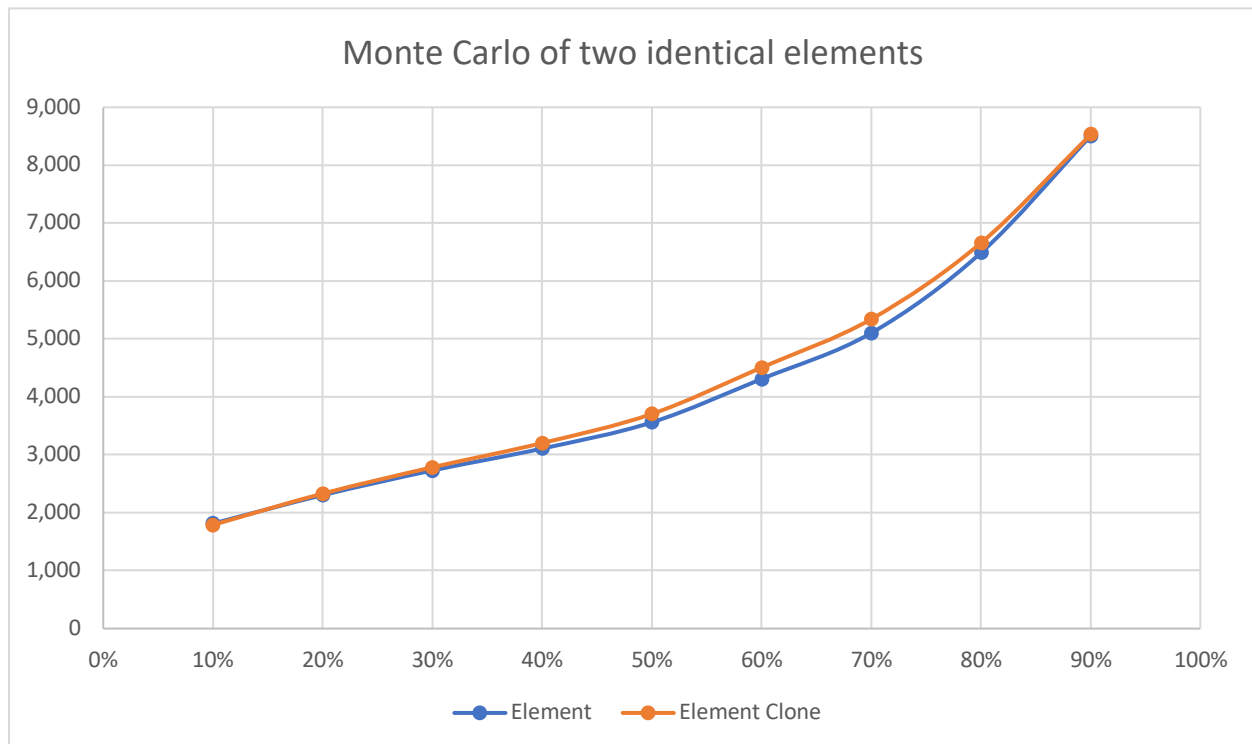

Figure 10. Monte Carlo of two identical elements

Monte Carlo in SEER from the User Perspective

A Monte Carlo run is performed when the user enables Monte Carlo Risk. The user can choose to modify their options by opening the Monte Carlo Risk Calc Options dialog from the drop-down on the Monte Carlo Risk button. This button can be found on the ribbons for report, chart, or options. The number of iterations can be set between 100 and 10,000. Correlation is specified

between WBS elements. SEER-H and SEER-SEM use a default of 60% of correlation while the rest of the SEER products (SEER-MFG, SEER-IT, SEER-SYS) set the option to fully uncorrelated. Figure 11 shows how you enable Monte Carlo in SEER-SEM.
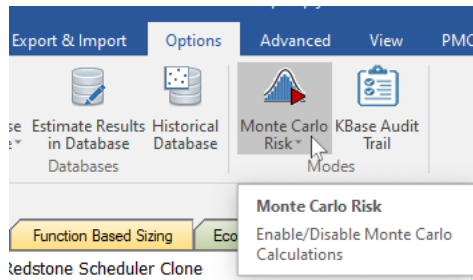

Figure 11. Monte Carlo can be enabled from the Report or Options menu

Once enabled, Monte Carlo will run when you recalculate (either change a parameter or F9). Monte Carlo risk can be run at the rollup level or at the element level. Enabling Monte Carlo will activate dedicated reports, and if you have risk charts turned on, those will utilize the results.

Monte Carlo Reporting. Results from the Monte Carlo can be accessed in report, chart or exportable data. An example of a Monte Carlo report is shown in Figure 12. There are three main sections:

> **Estimates by Confidence Levels**. Hours, Cost and/or Schedule are shown by **confidence level**. The confidence level can be interpreted as the probability that the actual result will be at or below the estimate. Think of it as a probability of not exceeding the estimate. (SEER-H has an option to provide Base Year vs Then Year cost.)

> **Summary Statistics and Assumptions**. The **mean** and **standard deviation** for the estimate. The coefficient of variation (CV) can be computed as the StdDev/Mean which is a relative measure of the spread. Assumptions on iterations and correlation are also included.

> **WBS Allocation by Confidence Level**. Selectable confidence level can be used to show how a rollup estimate will flow down into the elements. This is applicable at the rollup level only.
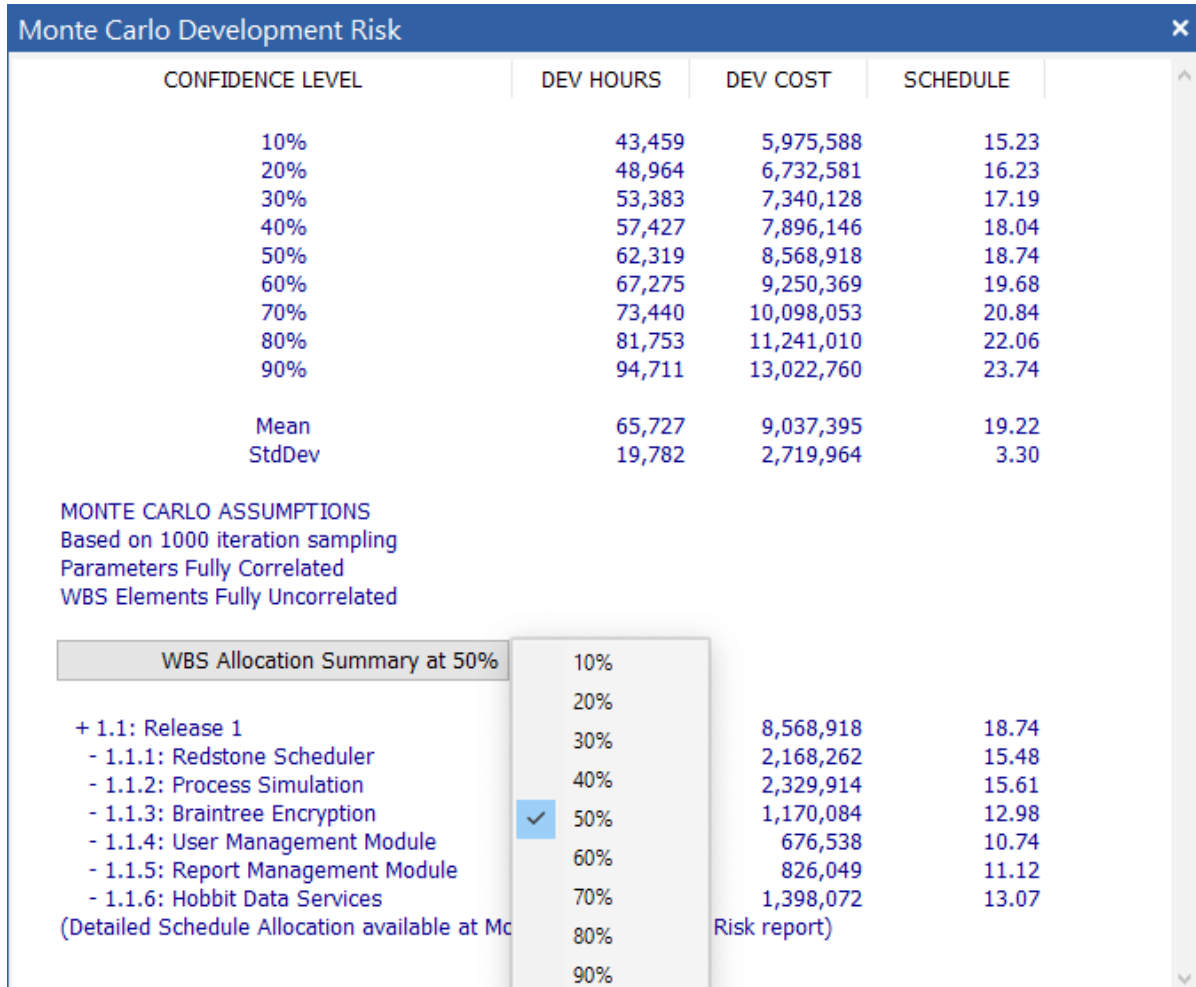
Figure 12. Monte Carlo report in SEER-SEM

Parameter Correlation

In the discussion thus far, the concept of parameter correlation has not been mentioned. In general, when SEER runs Monte Carlo, all parameters are correlated. That is, they all use the same probability when computing the estimate.   SEER-SEM has a unique feature that enables you to decouple the correlation between size and other parameters.   The concept here is that the things that drive size within a specified range may be different from parameters that drive productivity. If size is uncorrelated from other parameters, the estimate ranges will not be as extreme, especially on the high end.   Figure 13 shows the Monte Carlo options in SEER-SEM including the assumption used for correlating size and other parameters.
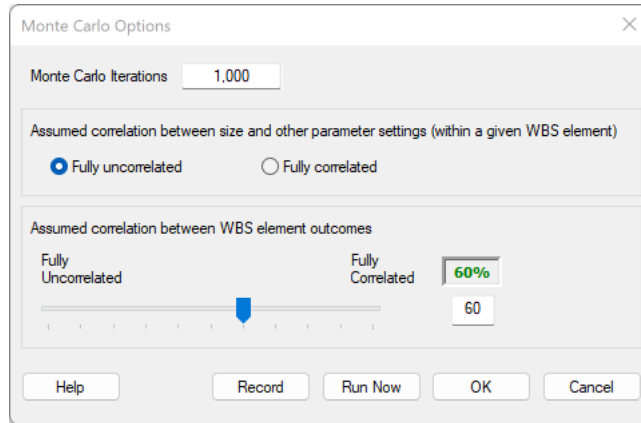
Figure 13. Monte Carlo options in SEER-SEM

Monte Carlo Options

Once enabled, you can view and set options for your Monte Carlo run. While these options vary a bit among the different SEER products, they generally are used to set the correlation and number of iterations.

If you are running SEER-H or SEER-MFG, you can choose which outputs to sample. This can help reduce processing time if you are not concerned with certain output categories. There is also an option to set the cost outputs to base year or then year cost. Figure 14 shows Monte Carlo options in SEER-H and SEER-MFG, respectively.
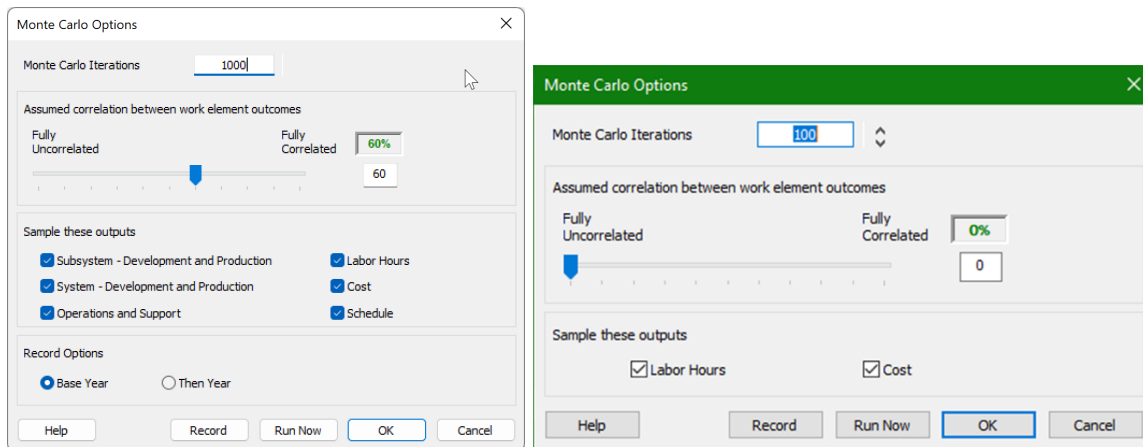

Figure 14. SEER-H and SEER-MFG Monte Carlo options

The record option will create a text file with the calculation for each iteration. This includes the random probability draw. This data can be used to check the math and help understand the nature of the Monte Carlo sampling.

The recorded file will have one row for each iteration that includes the probability used and the computed outputs. An example of the recorded data can be seen in Figure 15.

```
File    Edit    View

1/31/2023  4:54:12 PM    : WBS Description: Redstone Scheduler; Element Type: Program
1/31/2023  4:54:12 PM    : Parameters Fully Correlated
1/31/2023  4:54:12 PM    : WBS Elements 60% Correlated
1/31/2023  4:54:12 PM    : Iteration:      ProbSize:    ProbTech:   DevHrs:     DevCost:     DevSched:   MaintHrs:   MaintCost:
1/31/2023  4:54:12 PM    : 1    37.910000   37.910000   3082.656547 423865.275256    9.899339    0.000000    0.000000
1/31/2023  4:54:12 PM    : 2    5.010000    5.010000    1417.788744 194945.952292    7.610876    0.000000    0.000000
1/31/2023  4:54:12 PM    : 3    26.250000   26.250000   2580.619999 354835.249917    9.320852    0.000000    0.000000
1/31/2023  4:54:12 PM    : 4    70.090000   70.090000   5294.006532 727925.898136    11.941726   0.000000    0.000000
1/31/2023  4:54:12 PM    : 5    39.460000   39.460000   3149.621971 433073.021004    9.971682    0.000000    0.000000
1/31/2023  4:54:12 PM    : 6    77.060000   77.060000   6139.402107 844167.789694    12.576650   0.000000    0.000000
1/31/2023  4:54:12 PM    : 7    18.880000   18.880000   2246.594730 308906.775351    8.893484    0.000000    0.000000
1/31/2023  4:54:12 PM    : 8    11.120000   11.120000   1843.791549 253521.338055    8.318173    0.000000    0.000000
1/31/2023  4:54:12 PM    : 9    97.560000   97.560000   13630.73142 1874225.571505   16.608972   0.000000    0.000000
1/31/2023  4:54:12 PM    : 10   70.930000   70.930000   5384.950553 740430.701011    12.013089   0.000000    0.000000
1/31/2023  4:54:12 PM    : 11   97.230000   97.230000   13178.79328 1812084.076593   16.415284   0.000000    0.000000
1/31/2023  4:54:12 PM    : 12   58.890000   58.890000   4269.268460 587024.413197    11.075443   0.000000    0.000000
1/31/2023  4:54:12 PM    : 13   69.950000   69.950000   5279.089486 725874.804308    11.929944   0.000000    0.000000
1/31/2023  4:54:12 PM    : 14   50.350000   50.350000   3644.779499 501157.181171    10.478293   0.000000    0.000000
1/31/2023  4:54:12 PM    : 15   32.200000   32.200000   2837.718040 390186.230455    9.625546    0.000000    0.000000
1/31/2023  4:54:12 PM    : 16   83.190000   83.190000   7131.488826 980579.713585    13.252301   0.000000    0.000000
1/31/2023  4:54:12 PM    : 17   6.540000    6.540000    1542.662875 212116.145327    7.831251    0.000000    0.000000
1/31/2023  4:54:12 PM    : 18   54.570000   54.570000   3941.007011 541888.464075    10.769264   0.000000    0.000000
1/31/2023  4:54:12 PM    : 19   70.180000   70.180000   5303.633539 729249.611654    11.949319   0.000000    0.000000
1/31/2023  4:54:12 PM    : 20   65.610000   65.610000   4846.939951 666454.243286    11.578689   0.000000    0.000000
1/31/2023  4:54:12 PM    : 21   5.520000    5.520000    1461.515993 200958.448972    7.689449    0.000000    0.000000
1/31/2023  4:54:12 PM    : 22   3.210000    3.210000    1238.197865 170252.206389    7.270269    0.000000    0.000000
1/31/2023  4:54:12 PM    : 23   54.580000   54.580000   3941.735595 541988.644320    10.769981   0.000000    0.000000
1/31/2023  4:54:12 PM    : 24   51.920000   51.920000   3752.508506 515969.919630    10.585852   0.000000    0.000000
1/31/2023  4:54:12 PM    : 25   74.310000   74.310000   5778.926068 794602.334369    12.313420   0.000000    0.000000
1/31/2023  4:54:12 PM    : 26   76.820000   76.820000   6106.266505 839611.644478    12.552884   0.000000    0.000000
1/31/2023  4:54:12 PM    : 27   60.580000   60.580000   4405.975025 605821.565880    11.198390   0.000000    0.000000
1/31/2023  4:54:12 PM    : 28   73.360000   73.360000   5663.266685 778699.169155    12.226683   0.000000    0.000000
1/31/2023  4:54:12 PM    : 29   84.010000   84.010000   7291.754234 1002616.207171   13.355537   0.000000    0.000000
1/31/2023  4:54:12 PM    : 30   32.550000   32.550000   2852.717474 392248.652672    9.642750    0.000000    0.000000
1/31/2023  4:54:12 PM    : 31   10.120000   10.120000   1784.216638 245329.787691    8.226245    0.000000    0.000000
1/31/2023  4:54:12 PM    : 32   88.740000   88.740000   8424.115823 1158315.925611   14.045600   0.000000    0.000000
1/31/2023  4:54:12 PM    : 33   53.370000   53.370000   3854.554218 530001.204966    10.685886   0.000000    0.000000
```

Figure 15. Monte Carlo recorded details in a text file

Conclusion and Next Steps

Correlation is a critical component of cost risk analysis. Overlooking it leads to a significant underestimation of the total amount of uncertainty. However, for a large WBS, assigning correlation values for each individual pair of distinct elements can be daunting. The process can take a long time, as the number of correlations grows as the square of the size of the WBS and you have to be careful to ensure that the correlation values assigned are consistent. The events that have the biggest potential to cause cost to grow are things that will influence all WBS elements, such as schedule delays. Thus there is likely some amount of positive correlation among most, if not all, WBS elements. Taking this into account, a simple way to cut the proverbial Gordian knot of assigning correlation values is to assign a single value. Depending upon how you consider the problem, that single best value is somewhere in the range between 20% and 60%.

The SEER software products are a sophisticated suite of tools that provide a cost risk capability. Until now, however, this has been limited to either no correlation or 100% correlation across all WBS elements. The ability to choose a single value between 0% and 100% is now implemented in SEER-H, SEER-MFG and SEER-SEM, and will soon be in other SEER products. We have described how this works and how it affects the risk outputs from these models. Furthermore, Monte Carlo analysis can be performed on a combined SEER estimate (such as hardware and software) with consideration of correlation.

Some may argue that a single number is too simple. One way to make this more granular is to assign different values at WBS roll-up levels. This would allow, for example, modeling hardware structural, mechanical, and thermal elements with a single correlation value, while assigning a different value between those elements and avionics elements. Also, correlation is only one measure of stochastic dependency. There are others – in particular, tail dependency. Often, extreme risks occur together. Correlation does not capture this phenomenon, but tail dependency does. Both correlation and tail dependency can be modeled together. The use of tail dependency in risk modeling is discussed in an ICEAA paper from 2015 (Smart 2015).

References
1. S.A. Book, "Why Correlation Matters in Cost Estimating," Advanced Training Session, 32nd Annual DOD Cost Analysis Symposium, Williamsburg VA, 1999.
2. D. Mackenzie, "Cost Variance in Idealized Systems," International Society of Parametric Analysts Annual Conference, Cannes, France, 1996.
3. C.B. Smart, "Robust Default Correlation Values for Cost Risk Analysis," International Cost Estimating and Analysis Annual Conference, New Orleans LA, 2013.
4. C.B. Smart, "Beyond Correlation: Don't Use the Formula That Killed Wall Street," International Cost Estimating and Analysis Annual Conference, San Diego CA, 2015.
5. C.B. Smart, *Solving for Project Risk Management: Understanding the Critical Role of Uncertainty in Project Management*. New York McGraw-Hill 2020.