Spread Too Thin

Managing Coefficient of Variation in Monte-Carlo Based Cost Models

Stephen Koellner

Abstract

Coefficient of Variation (CV) can be utilized to determine whether sufficient uncertainty is captured in Monte-Carlo based estimates. This topic explores common barriers to capturing program level risk using the interpretation of a WBS as a linear combination of distributions. A WBS CV equation is provided to model perturbations of a baseline case and then randomized WBSs are generated to analyze CV at scale. Estimators can apply these insights to improve program estimate risk calculations.

ICEAA Professional Development & Training Workshop

2023-02-21

Table of Contents

Table of Contents 2

1. Introduction 3
   - 1.1. The Cost Estimating Process 3
   - 1.2. Work Breakdown Structures 4
2. Calculating Cost Uncertainty 6
   - 2.1. Top-Down Application of Cost Uncertainty 7
   - 2.2. Bottom-Up Application of Cost Uncertainty 8
   - 2.3. Top-Down vs Bottom-Up Application 10
   - 2.4. Coefficient of Variation (CV) 10
3. Interpretation of a WBS as a Convolution of Distributions 11
   - 3.1. Correlation Discussion 13
   - 3.2. Example Calculation & Monte-Carlo Simulation Result 14
4. Behavior of CV Equation 15
   - 4.1. Perturbations of a Baseline Case 16
   - 4.2. Summary of Perturbations and Extreme Case 17
   - 4.3. Over Sharpening the Pencil 18
5. Randomized WBS 19
   - 5.1. Process of Randomized WBS 19
   - 5.2. Sampling Statistics of Subordinate WBS Elements 20
   - 5.3. Large-Scale Results of Simulations 22
   - 5.4. WBS Size Rule of Thumb 24
   - 5.5. Implications for Project Management 24
6. Conclusion 25

References 26

Appendix A: Monte-Carlo Based Cost Modeling 27

Appendix B: Special Case of Independence 29

Appendix C: Perturbations to Baseline Case 31

AuGUR CONSULTING

1. Introduction

The Government Accountability Office (GAO) Cost Estimating and Assessment Guide states that: "A credible estimate includes a risk and uncertainty analysis that quantifies the imperfectly understood risks and identifies the effects of changing key cost driver assumptions and factors." [1, p. 33] All cost estimates need to quantify reasonable cost growth to minimize future funding risks, determine appropriate variances of realized costs, and to calculate sufficient values for management reserve and contingency. Cost estimators develop and present the output of a cost model as a probability distribution with a range of potential future costs, rather than just a single number. Program leadership and decision-makers require a tangible snapshot of this output distribution as well as an overall level of future cost variances to make fully informed decisions.

Quantifying cost uncertainty using a bottom-up application of cost uncertainty has the benefit of being able to represent output spread as a function of input variable distributions, which in-turn can be utilized to conduct risk-based cost driver analysis. A significant drawback of this method is that output-level cost uncertainty can be easily underestimated unless appropriate measures are taken to counteract the impacts of an overly robust WBS or insufficient accounting of correlation among input variables. These impacts can be shrouded from an analyst, especially when using a Monte-Carlo based simulation to approximate these interactions in a cost model. It is not uncommon for a cost estimator to review risk-adjusted cost output of a model and observe a top-level spread that does not easily trace to the spread of lower level cost elements and input variables.

Understanding how modeling choices impact output spread can minimize the likelihood of funding risks during program execution or false funding risk actualization if actual expenditures exceed the value of unrealistic upper cost bounds. This paper seeks to illustrate the relationship between input and output distributions by defining output CV as a function of the children level elements in a WBS and using this equation to provide cost estimators with guidance to prevent overoptimistic projections within their cost models. Additionally, conditions for WBS size across the acquisition lifecycle are provided alongside a rule of thumb to identify whether an estimate is overly optimistic.

1.1. The Cost Estimating Process

Presented below in Figure 1 is an iterative cost estimating process that is adapted from GAO best practices of cost estimation [1, p. 34]. Each step ensures that the resulting cost estimate fulfills its purpose, is based on consistent and repeatable methodology, and provides meaningful information and insight to programmatic decision-makers.

Mathematical models, including dynamic cost estimates, are used to simulate the relationships and behavior between input variables and modeled outputs to better understand a system and predict actual outcomes. Intricate models of even relatively simple systems can provide a greater level of insight than simpler models, provided that the proper steps are taken to create an accurate model and interpret the results. Cost models that do not account for risk and uncertainty are simple to develop but can be limited in terms of aiding programmatic decision making. Fully fledged cost models incorporating statistical uncertainty are more intensive to develop but can provide a plethora of insight which can enable cost-risk reducing decision making from program leadership.

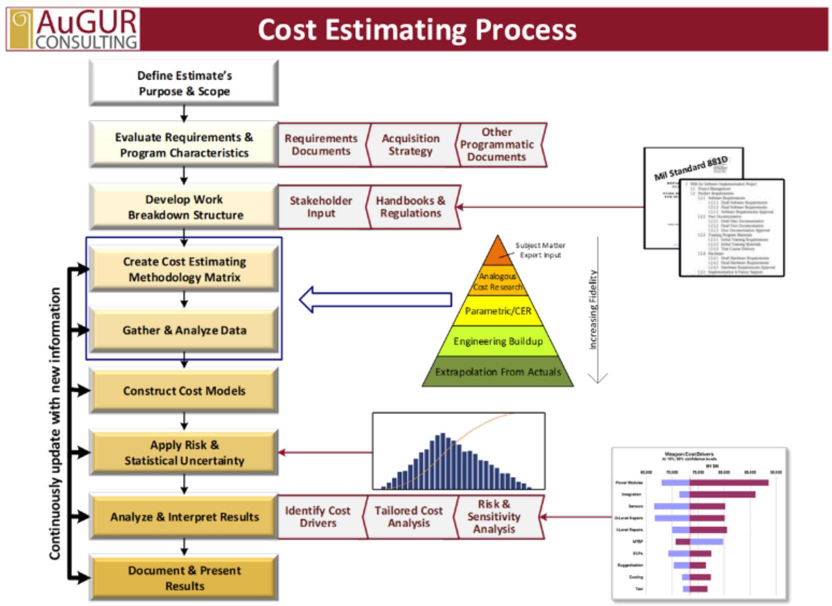


Figure 1 – The Cost Estimating Process

A key step to identifying these insights is to analyze and interpret the results of a model thoroughly, including input variable sensitivity and detailed uncertainty analysis. A drawback of intensive modeling is that the complex nature of detailed modeling can obscure the behavior of certain input variables and make these insights more difficult to uncover. It is imperative that an analyst understands all aspects of their modeling techniques to identify these insights, and it a goal of this paper to explain recurring phenomena of detailed cost models so cost estimators can more easily identify insights relevant to program leaders.

1.2. Work Breakdown Structures

A work breakdown structure (WBS) displays and defines the product(s) to be developed and produced [2, p. 1]. Within the context of program management and cost estimation, the WBS is the organizational construct that decomposes and relates all elements of a program. All aspects of a program such as Systems Engineering, Training, System Test and Evaluation, Industrial Spares and Repair Parts, etc., can all be captured within a WBS. Subordinate elements are indented below their parent element in the WBS, and the number of indentures on an element refers to the WBS Level of that element. Parent level elements of a WBS in a cost model are exactly the sum of the costs of the subordinate WBS elements below it.

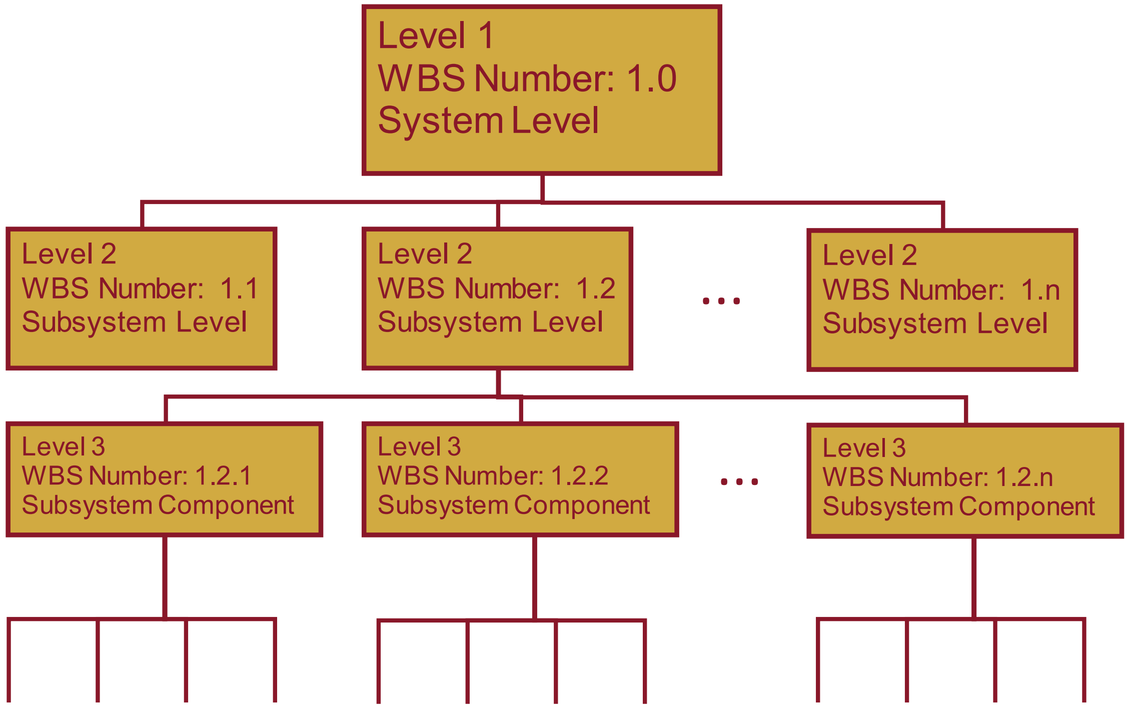


Figure 2 – Notional WBS

This paper explores the translation of cost uncertainty inside of a WBS, and it is critical to first define what a WBS is in the context of cost estimation before defining how cost uncertainty is incorporated within a cost model. For cost estimates, the WBS organizes and displays the individual costs elements of the estimate. For example, Table 1 is a depiction of a WBS for a contract cost estimate broken down to the level two elements. All the costs associated with the level two elements sum to the total cost at the level one element. Each one of the level two elements can either be the point where an independent calculation is made at (referred to as a lowest level element) or broken even further down into level three elements. For instance, management costs can be estimated as a total headcount multiplied by a weighted average labor rate. Conversely, management costs can include level three elements below for each individual labor category/rate and even include non-labor costs.

WBS Level	WBS Element
1	**Total Contract Cost**
2	Management
2	Development Labor
2	Prototype Materials
2	Equipment
2	Testing Labor
2	Testing Equipment

Table 1 – WBS Decomposition

The summation of costs to the level one element of the WBS is the total value of the cost estimate, where each element in that sum is a subordinate cost element. This approach is adopted for two primary reasons:

1. The smallest, appropriate elements of the estimate can be identified and estimated.
2. Intermediate calculations of cost or range of potential cost outcomes for an element at any WBS level can be shown.

2. Calculating Cost Uncertainty

All cost estimates should account for cost uncertainty in some form, to mitigate the risks of cost variability and future project scope changes. Variances are to be expected between projected costs in early-stage estimates and actual costs, due to elements such as choice of vendor, productivity levels, or changing market conditions. Additionally, scope changes to a project plan will cause variances should additional work be required, quantity of systems or materials changes, operational plans evolve, etc. A wide range of changes to project elements should be anticipated between early-stage cost estimates and execution, and cost estimators must account for these potential changes within their cost models.

Cost estimators should build and interpret a cost model as a range or a distribution of different values rather than a single number, even though cost output is typically shown as a single, static number for practicality in funding requests and early-stage planning of a project. There are two primary methods for adding statistically uncertainty into a cost model: top-down application and bottom-up application. Each method has its strengths and weaknesses, and a cost modeler must determine which method is most appropriate based on the maturity of the estimate and agency/department guidance of the project. The central focus of this study is on the behavior of cost estimates utilizing the bottom-up application method. However, it is important to define and compare both application methods to understand the content and recommendations of this paper.

2.1. Top-Down Application of Cost Uncertainty

The first method shown for applying statistical uncertainty, is the top-down application. With the top-down application method, a cost estimator builds a point estimate and assigns a probability distribution at a total or near-total level WBS element that includes multiple cost elements and variables. The cost model point estimate is built using the most likely value of variables in a bottom-up cost estimate, and the cost estimator then assumes that the sum of the point estimate at the top of the WBS is also the most likely value. With this assumption, the cost estimator can then assign a probability distribution to the total point estimate, where the total point estimate itself serves as the peak of that probability distribution.

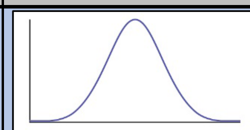
Top - Down Application of Cost Uncertainty			
WBS Level	**WBS Element**	**Equation**	**Application of Risk**
1	**Total Contract Cost**	*Sum of Children*	
2	Management	*Labor Pool 1 + Labor Pool 2 + Labor Pool 3*	*None*
2	Development Labor	*Headcount x Labor Rate*	*None*
2	Prototype Materials	*Quantity * Unit Cost*	*None*
2	Equipment	*Base Cost + Complexity Factor*^{2}	*None*
2	Testing Labor	*Historical Cost + (Factor x Test Quantity*^{2}*)*	*None*
2	Testing Equipment	*Equipment 1 + Equipment 2*	*None*

Table 2 – Top-Down Application of Cost Uncertainty

The choice of the assigned probability distribution is based on either actual cost data from analogous projects at the total or near-total level, or on industry guidance when data is too limited. This method of applying cost uncertainty to a model has the inherent advantage of data availability for choosing distributions, as there are usually more data points available for high-level cost elements rather than individual lower-level cost input variables. It is also easier to adjust this distribution for specific risk factors at the higher levels of a WBS since this method reduces the number of distributions required to implement. Additionally, this method is computationally trivial to implement and thus reduces the time to run calculations and draw insights from the cost outputs.

The top-down application method, however, does have specific weaknesses which limit its usefulness to cost estimators and the project leadership that they report to. There is limited insight to specific cost input drivers from a statistical standpoint, as all probability distributions are assigned at the cost output level rather than to individual inputs. A highly volatile cost input variable may have significant impacts to uncertainty at the total cost level, but it would not be observable outside of pure model sensitivity when using this method of uncertainty application. Since the assumption of cost spread is only made at the top of the WBS, there are limited ways to portion out cost uncertainty to lower level WBS elements to determine which elements of the project are high risk. The assumption of top-level spread can also be interpreted as a weakness of this approach, as the distribution of top-level cost elements may be based on other projects which are not analogous to the project being estimated.

The culmination of strengths and weaknesses of this cost uncertainty application method necessitates that this method should only be utilized for very early-stage cost estimates, rough order of magnitude cost estimates, or when specific agency guidance enforces this method. In all other situations, the bottom-up application method should be the preferred choice of the estimator.

2.2. Bottom-Up Application of Cost Uncertainty

The bottom-up application method can be thought of as a refinement of cost modeling techniques used in the top-down application method, but this refinement requires more information, time, and effort to correctly implement. Like the top-down method, the cost estimator builds a bottom-up cost model to a point estimate value, but rather than applying a probability distribution at the total or near-total element(s) the estimator assigns probability distributions to individual cost input variables. This method shifts the assumption of cost spread away from total or near-total cost elements in the WBS and towards specific input variables such as labor rates, unit costs, percentage factors, etc. By assigning probability distributions to input variables and accounting for sufficient correlation of input variables, an estimator can create a more tailor-made cost model that can be used to identify high risk elements that are unique to that specific program. Insights to specific cost drivers, such as individual labor pools, subcomponents of a systems, software license renewals, etc., can be derived from a model utilizing a bottom-up application of uncertainty. Additionally, relationships between requirements and the resulting cost uncertainty can be observed with this modeling technique.

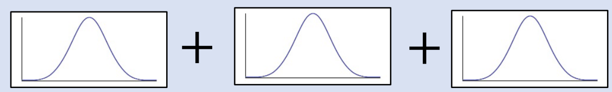
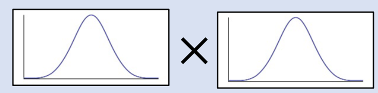
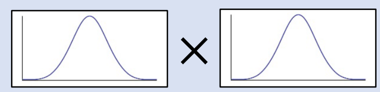
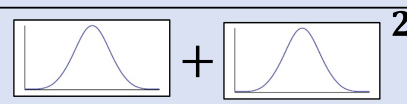
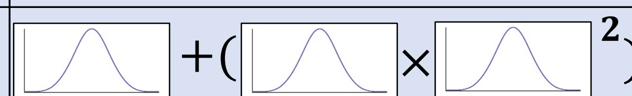
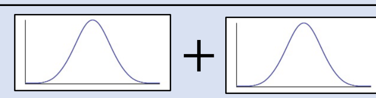
Bottom - Up Application of Cost Uncertainty			
WBS Level	**WBS Element**	**Equation**	**Application of Risk**
1	**Total Contract Cost**	*Sum of Children*	*Composition of Children*
2	Management	*Labor Pool 1 + Labor Pool 2 + Labor Pool 3*	
2	Development Labor	*Headcount x Labor Rate*	
2	Prototype Materials	*Quantity * Unit Cost*	
2	Equipment	*Base Cost + Complexity Factor*^{2}	
2	Testing Labor	*Historical Cost + (Factor x Test Quantity*^{2}*)*	
2	Testing Equipment	*Equipment 1 + Equipment 2*	

Table 3 – Bottom-Up Application of Cost Uncertainty

The bottom-up application method also has some drawbacks which can limit its viability for certain estimates. Applying cost uncertainty at the lowest level of an estimate requires substantially more time and data to implement correctly, and more assumptions about the distributions of input variables may be necessary to complete an estimate. Numerical calculations of this method can be computationally taxing when using deterministic methods, and so heuristic methods such as a Monte-Carlo simulation (see Appendix A) are necessary to generate output in practice. In addition to requiring more effort to develop, fully analyzing, and interpreting the results of this type of model is more intensive than a top-down application method. Lastly, the bottom-up application of uncertainty is prone to underestimating cost spread at higher levels unless done correctly, as will be explored in this paper.

Despite requiring more effort and care to correctly implement, the bottom-up application of cost uncertainty provides the opportunity for unique cost and risk insights to be made about a program that the top-down application method cannot provide. As such, cost estimates which are used to support requirements documents, annual funding requests, major contract negotiations, or major decision points of a program should follow the bottom-up application method of cost uncertainty.

2.3. Top-Down vs Bottom-Up Application

Comparing Application Methods of Cost Uncertainty		
Application	**Pros**	**Cons**
Top - Down	• Simplifies cost modeling • Generally, more data is available to defend top level spread	• Limited ability to analyze cost drivers and quantify impact to model spread • Assumptions on spread not directly traceable to inputs Range of cost outcomes can only be viewed at top-level
Bottom - Up	• Spread of total cost directly depends on cost inputs • Range of cost outcomes can be viewed for any WBS element	• Complicates cost modeling/behavior of cost model • Can more easily underestimate cost uncertainty

Table 4 – Comparing Applications of Uncertainty

The previous sections outlined the strengths and weaknesses of each application method. Cost estimators must choose the method most appropriate for the estimate based on several different factors, including:

1. Agency/department guidance on modeling cost uncertainty.
2. The purpose and scope of the cost estimate.
3. Availability of data.
4. Maturity of the program or project being estimated.
5. Time and resources available to the cost estimator.

Cost estimators must evaluate the characteristics of each application method against the above criteria and determine which method is appropriate. Ideally, multiple cost models developed by independent estimators with varying methodology and uncertainty application methods can be used to compare results to develop a consensus output, but this is often not practical or viable in most programs. It is common to see only one method used in a cost model, and when only application can be implemented it is desirable to use the bottom-up application method for the many benefits afforded by that method.

2.4. Coefficient of Variation (CV)

In cost estimating, risk-adjusted models following the bottom-up application method produce a distribution of total cost by approximating the sum of distributions corresponding to cost elements that compose the total cost, via a Monte-Carlo simulation. The CV is a significant statistic to utilize when interpreting and analyzing the results of a risk-adjusted cost model, as it "normalizes" the standard deviation of distributions as a percentage of the mean. This "normalization" allows for comparison between different risk-adjusted cost elements irrespective of their magnitudes, providing a single metric describing the relative degree of uncertainty on a cost element. The CV can also be used to compare cost spread between lower-level elements within a single model, and relatively higher risk aspects of a program can be identified, assessed, and potentially mitigated through program management actions.

For a probability distribution X with standard deviation σ_X and mean μ_X , the CV (sometimes referred to as the coefficient of dispersion [1, p. 94]) of X is defined as:

$$CV_X := \frac{\sigma_X}{\mu_X} \quad (1)$$

Where $\mu_X = E[X]$ [1, p. 76] and ${\sigma_X}^2 = E[(X - \mu_X)^2]$ [1, p. 77]. This definition itself is not useful for analyzing the top-level results of a cost model, as calculating the CV requires that the standard deviation and mean are known values. In practice, a cost estimator would apply probability distributions to individual input variables of a model or cost elements in a WBS, and then run a Monte-Carlo simulation to estimate the standard deviation and the mean of the results. It can sometimes be difficult to interpret the top-level CV, especially in large cost models where the complicated interactions between input variables obfuscate the mechanics of the resulting probability distributions.

If a cost estimator took a cost model and wrote out an equation for the total cost of a large program by hand, down to the lowest appropriate input variable, the resulting equation could fill up multiple pages. Assigning probability distributions to input variables and then attempting to determine the resulting probability distribution of the total cost would be an extremely time-consuming, cumbersome, and mathematically challenging endeavor. Monte-Carlo simulations allow for these calculations to be approximated in minutes, but the underlying complexity of these calculations remain, and further analysis of the results requires an understanding of these mechanics.

3. Interpretation of a WBS as a Convolution of Distributions

In probability theory, a linear combination of random variables is referred to as a convolution of probability distributions. If random variables A and B are summed together to form C, then C is called the convolution of random variables A and B. [1, p. 219] The WBS of a risk-adjusted cost model developed under the bottom-up application of cost uncertainty can also be thought of as a convolution of the probability distributions of all the lowest-level elements in the WBS. In this interpretation, a universal relationship between the uncertainty of lower-level elements and the resulting uncertainty of the top-level element can be derived.

WBS Level	WBS Element
1	Total Contract Cost
2	Management
2	Development Labor
2	Prototype Materials
2	Equipment
2	Testing Labor
2	Testing Equipment

WBS Level	WBS Element
k	Z
k+1	X_1
k+1	X_2
k+1	.
k+1	.
k+1	.
k+1	X_n

Figure 3 – WBS as a Convolution of Probability Distributions

Let Z be a WBS element of a cost model developed using the bottom-up uncertainty application method, with n lowest level WBS subordinate elements X_i. Each X_i is a random variable since this is a risk-adjusted cost model, and the following convolution is the definition of random variable Z.

$$Z = \sum_{i=1}^{n} X_i \qquad (2)$$

The goal of this interpretation is to define the CV of Z as a function of the statistics of its subordinate elements X_i. Start with substituting the definition of Z shown in equation (2) into the definition of CV shown in equation (1):

$$CV_Z = \frac{\sigma_Z}{\mu_Z} = \frac{\sqrt{Var(Z)}}{E[Z]} = \frac{\sqrt{Var(\sum_{i=1}^{n} X_i)}}{E[\sum_{i=1}^{n} X_i]} \qquad (3)$$

Since linearity is preserved in the expected value function [1, p. 76], the denominator of equation (3) becomes:

$$\frac{\sqrt{Var(\sum_{i=1}^{n} X_i)}}{E[\sum_{i=1}^{n} X_i]} = \frac{\sqrt{Var(\sum_{i=1}^{n} X_i)}}{\sum_{i=1}^{n} E[X_i]} \qquad (4)$$

In the numerator of equation (4), the variance equation needs to be expanded using the definition of variance and covariance (COV) [1, p. 172]:

$$Var\left(\sum_{i=1}^{n} X_i\right) = E\left[\left(\sum_{i=1}^{n} X_i\right)^2\right] - E\left[\left(\sum_{i=1}^{n} X_i\right)\right]^2 = E\left[\sum_{i=1}^{n}\sum_{j=1}^{n} X_i X_j\right] - \sum_{i=1}^{n}\sum_{j=1}^{n} E[X_i]\,E[X_j]$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} E[X_i X_j] - E[X_i]E[X_j] = \sum_{i=1}^{n}\sum_{j=1}^{n} COV(X_i, X_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} r_{i,j}\sqrt{Var(X_i)Var(X_j)} \quad (5)$$

For any i, j the covariance can be expressed as $COV(X_i, X_j) = r_{i,j}\sqrt{Var(X_i)\,Var(X_j)}$ where $r_{i,j}$ is the Pearson correlation coefficient between random variables X_i and X_j. [1, p. 173] Substituting equation (5) into the numerator of equation (4) results in the following relationship:

$$CV_Z = \frac{\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} r_{i,j}\sqrt{Var(X_i)\,Var(X_j)}}}{\sum_{i=1}^{n} E[X_i]} = \frac{\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} r_{i,j}\,\sigma_{X_i}\,\sigma_{X_j}}}{\sum_{i=1}^{n}\mu_{X_i}} \quad (6)$$

Lastly, the definition of CV in equation (1) for each random variable X_i can be substituted into equation (6) to derive the result:

$$CV_Z = \frac{\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} r_{i,j}\,(CV_{X_i}\,\mu_{X_i})\,(CV_{X_j}\,\mu_{X_j})}}{\sum_{i=1}^{n}\mu_{X_i}} \quad (7)$$

The result in equation (7) is valid for any bottom-up application method where the top-level element Z is a convolution of subordinate WBS elements that are representative of random variables X_i. Note that this relationship is not dependent on the choice of distribution, shape, or skew of the random variables X_i.

3.1. Correlation Discussion

A crucial aspect of the derived equation is the inclusion of correlation between elements in the WBS, which was not an immediate area of consideration at the beginning of this study. When implementing a bottom-up approach to applying cost uncertainty, an estimator should be defining correlation groups between relevant cost input variables to account for. These correlation groups at the input level will typically have an impact at the WBS and will translate to non-zero correlation between the resulting probability distributions of lowest level WBS elements. Additionally, dynamic models will commonly re-use input variables at various points in the WBS. For example, a variable for system unit cost could be used to calculate the prime mission product as well as to calculate the replacement costs of additional systems. This cross-utilization of input variables across the WBS will result in functional correlation. Monte-Carlo simulations will account for both applied and functional correlation between input variables, but it is important to understand the presence of correlation in the aggregate to fully interpret the cost uncertainty of the outputs.

For further discussion on the impact of independence between WBS elements, see Appendix B and Appendix C.

3.2. Example Calculation & Monte-Carlo Simulation Result

To further verify the validity of the CV equation, it remains to be demonstrated the comparison of a CV calculated using the equation and the CV produced from a Monte-Carlo simulation. Since a Monte-Carlo simulation approximates a probability distribution, rather than calculating it exactly, a small but non-zero difference between the CV equation and the result of the simulation will be sufficient to demonstrate the validity of the derived CV equation. It should be noted that in general, this delta should decrease as the number of iterations in the simulation are increased.

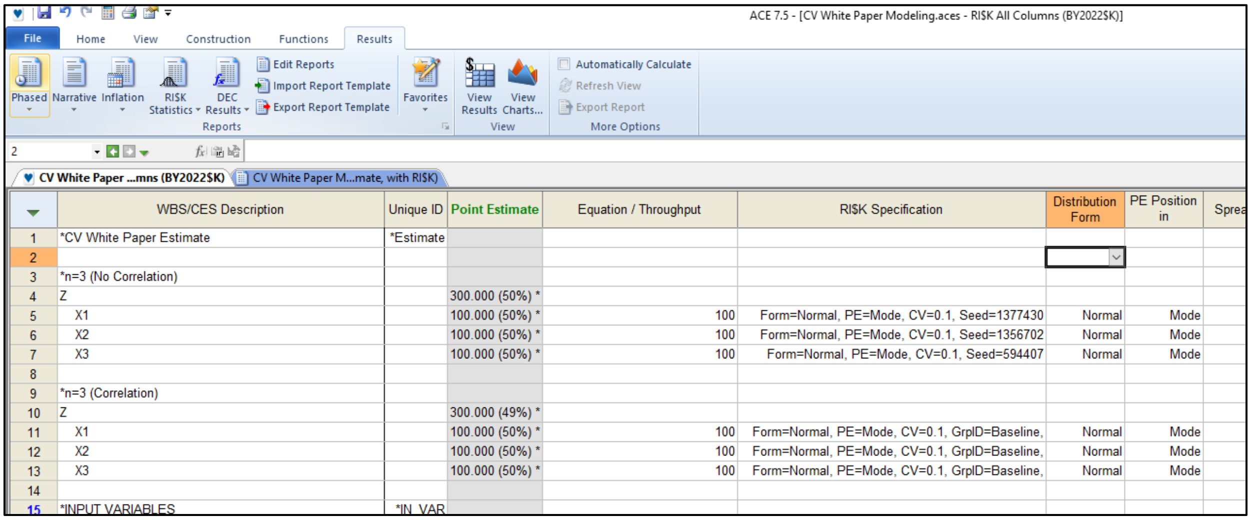
The demonstration will utilize a simple WBS with three identical elements being summed together, along with a correlation matrix to capture the correlation between WBS elements. A calculation of the top-level element's CV will be done using the derived CV equation, and the result will be compared to the CV calculated from running a Monte-Carlo simulation under the same conditions. Tecolote's Automated Cost Estimator (ACE) 7.5 application will be used to run the Monte-Carlo simulation as it allows for correlation matrices to be defined explicitly.

Simplified WBS (n=3)			
WBS	**Mean**	**Stan. Dev.**	**CV**
Z	**300**		
X1	100	10	0.1
X2	100	10	0.1
X3	100	10	0.1

Correlation Matrix			
	X1	**X2**	**X3**
X1	1	0.25	0.25
X2	0.25	1	0.25
X3	0.25	0.25	1

Figure 4 – Simple WBS for CV Calculation Example

The WBS is comprised of three identical elements, each with a mean value of 100 and a CV of 0.1. The correlation matrix is comprised of values of 0.25 for the off-diagonal entries, and the diagonal values are, by definition, set to one.



	WBS/CES Description	Unique ID	Point Estimate	Equation / Throughput	RISK Specification	Distribution Form	PE Position in	Sprea
1	*CV White Paper Estimate	*Estimate						
2								
3	*n=3 (No Correlation)							
4	Z		300.000 (50%) *					
5	X1		100.000 (50%) *	100	Form=Normal, PE=Mode, CV=0.1, Seed=1377430	Normal	Mode	
6	X2		100.000 (50%) *	100	Form=Normal, PE=Mode, CV=0.1, Seed=1356702	Normal	Mode	
7	X3		100.000 (50%) *	100	Form=Normal, PE=Mode, CV=0.1, Seed=594407	Normal	Mode	
8								
9	*n=3 (Correlation)							
10	Z		300.000 (49%) *					
11	X1		100.000 (50%) *	100	Form=Normal, PE=Mode, CV=0.1, GrpID=Baseline,	Normal	Mode	
12	X2		100.000 (50%) *	100	Form=Normal, PE=Mode, CV=0.1, GrpID=Baseline,	Normal	Mode	
13	X3		100.000 (50%) *	100	Form=Normal, PE=Mode, CV=0.1, GrpID=Baseline,	Normal	Mode	
14								
15	*INPUT VARIABLES	*IN_VAR						

Figure 5 – Simple WBS Calculation (Created using ACE, Version 7.5)

When using the derived CV equation, which is referred to as the True CV in the table below, the CV of the top-level element Z comes out to 0.0707. The same exact conditions modeled in ACE result in a CV of 0.0710 after 10,000 iterations of the Monte-Carlo simulation. This represents a -0.41% difference between the two values, further validating the derived equation. Increasing the number of iterations will generally cause the percent difference between the calculated value and the output from the Monte-Carlo simulation to converge to zero.

CV Calculation	
True CV (Eqn)	0.0707
Monte Carlo Sim*	0.0710
% Difference	-0.41%

** 10,000 Iterations*

Table 5 – Calculated CV Against Simulated CV

4. Behavior of CV Equation

The purpose of deriving the CV equation is to understand what elements of a cost model can impact the overall spread of cost outputs at the total or near total level. There are four main elements at the WBS which can impact the top-level CV:

1. The CV of the lowest level WBS elements
2. The mean of the lowest level WBS elements
3. The total number of lowest level elements in the WBS
4. The correlation between lowest level WBS elements

Cost estimators may encounter output CVs that do not align with their qualitative expectations of cost uncertainty in a model. This section will explore the behavior of the CV equation against these four main elements.

4.1. Perturbations of a Baseline Case

A simple WBS is established as the baseline case, and the resulting CV will serve as a benchmark against perturbations to the baseline case. Each perturbation will alter a single element of the baseline case, and the resulting CV of the perturbation will be compared to the benchmark CV to determine how each of the four main elements impact the level of cost variability in a WBS. The baseline case will again be comprised of three identical WBS elements, with means of 100, CVs of 0.1, and off-diagonal correlation coefficients of 0.25. The table below establishes the baseline case.

Simplified WBS (n=3)			
WBS	Mean	Stan. Dev.	CV
Z	**300**		
X1	100	10	0.1
X2	100	10	0.1
X3	100	10	0.1

Correlation Matrix			
	X1	**X2**	**X3**
X1	1	0.25	0.25
X2	0.25	1	0.25
X3	0.25	0.25	1

Table 6 – Baseline WBS Case

For perturbations to this baseline case, the mean, the CV, and the correlation coefficients of the lowest level WBS elements will remain constant unless that aspect of the model is being examined. Detailed output of all perturbations are shown in Appendix C, and the summarized output of perturbations are shown in Section 4.2.

AuGUR CONSULTING

4.2. Summary of Perturbations and Extreme Case

Behavior of CV			
Scenario	**CV**	**% Δ to Baseline**	**Note**
Baseline	**0.0707**	**0%**	***n=3, μ = 100, CV = 0.1, r = 0.25***
High Stan. Dev.	0.0972	37%	*Double one standard dev.*
Large Mean	0.0729	3%	*Double one mean*
Large Mean (Normalized)	0.0791	12%	*Double one mean, reduce mean of other elements*
Large WBS	0.0612	-13%	*Double WBS/maintain top-level mean*
Strong Correlation	0.0745	5%	*Double single correlation coefficient*
No Correlation	0.0577	-18%	*Model independent distributions*

Table 7 – Summary of Perturbations

Table 7 above summarizes all the examined perturbations to the baseline case, with all four aspects of the CV equation being adjusted across all the perturbations. The most impactful change to the baseline case was the increase to the standard deviation of a single element, and while that is an intuitive result that is not an adjustment to the cost model that an estimator can make without supporting data and rationale. Similarly, increasing the mean of a single WBS element will also increase the top-level CV, but this artificially increases the mean of the top-level element of the WBS as well. That leaves the grouping characteristics of lowest level WBS elements, the level of correlation between lowest level WBS elements, and the level of detail in the WBS as aspects of a model that can reasonably be altered without compromising the integrity of the model.

The following scenario compares two extreme cases that aggregate the positive and negative model attributes which do not impact the integrity of a risk adjusted cost model. The High-Spread WBS has grouped together elements in the first WBS element and a high level of correlation has been applied to all the WBS elements. The Low-Spread WBS has a six lowest level WBS elements, each of which are independent of one another. It is important to note that both scenarios have the same mean sum at the top-level as well as the same average CV of each lowest level WBS element. This comparison can be interpreted as two different cost models of the same project and scope.

High Spread WBS			
WBS	Mean	Stan. Dev.	CV
Z	300		
X1	200	20	0.1
X2	50	5	0.1
X3	50	5	0.1

Low Spread WBS			
WBS	Mean	Stan. Dev.	CV
Z	300		
X1	50	5	0.1
X2	50	5	0.1
X3	50	5	0.1
X4	50	5	0.1
X5	50	5	0.1
X6	50	5	0.1

CV Calculation	
High Spread	0.086603
Low Spread	0.040825
% Δ CV	-53%

Correlation Matrix			
	X1	X2	X3
X1	1	0.5	0.5
X2	0.5	1	0.5
X3	0.5	0.5	1

Correlation Matrix				
	X1	X2	...	X6
X1	1	0	0	0
X2	0	1	0	0
...	0	0	1	0
X6	0	0	0	1

Table 8 – High-Spread WBS & Low-Spread WBS Comparison

The High-Spread WBS results in a top-level CV of 0.0866 while the Low-Spread WBS has a top-level CV of 0.0408, which is a 53% difference from the High-Spread WBS. This scenario illustrates that two separate models with the same top-level mean value, can have drastically different levels of top-level cost uncertainty because of seemingly minor modeling techniques.

The findings of this study can be generalized to real world applications of diagnosing underestimated or overestimated top-level cost spread of risk-adjusted cost models. When determining why a cost model has a counter-intuitive cost spread, the following potential questions should be addressed:

- Is there an appropriate level of cost uncertainty at the input level of the model?
- Is there a sufficient level of correlation between input variables in the model?
- Are lowest level WBS elements grouped or consolidated appropriately?
- Does the level of detail in the WBS align with the qualitative interpretation of program certainty?

The above questions will either be answered in the affirmative or will require an adjustment to the model to ensure that cost uncertainty is not underestimated. All corrective actions done to a model must follow best practices and adhere to available data. Corrective actions could include items such as a re-evaluation of input level probability distributions, an increase or decrease to correlation coefficients, different groupings of elements in a correlation group, consolidation of elements in the WBS, or a complete overhaul to the level of complexity in the model.

4.3. Over Sharpening the Pencil

In cost modeling, there is a phenomenon called "Over Sharpening the Pencil", where an estimate is developed to a high level of fidelity and detail that does not align to the actual certainty of the scope of that estimate. An example of this would be an in-depth estimate for the production unit cost of an airplane while the project is still in the early developmental phase of its lifecycle. While having an estimate that accounts for the number of man-hours per component, the precise amount of raw materials required, and an engineering build-up that goes down to the number of screws looks accurate on paper, in reality that estimate injects a high-level of certainty for an item that is still being designed and tested. A more appropriate approach to this scenario is to rely on subsystem level build-up that leverages cost estimating relationships and analogies to other aircrafts. An overly precise estimate will lower the cost spread at the total or near-total level, as the number of lowest level WBS elements increases to account for the higher level of detail in the estimate. The perturbation for the larger WBS size indicates this behavior, and the following section will model this phenomenon.

5. Randomized WBS

The final goal of this study is to provide conditions for WBS size that will limit the likelihood of underestimated cost uncertainty in a model, assuming that all other best practices for cost modeling are followed. This goal is achieved through randomized simulations of typical cost estimator practices when developed a cost estimate. This will allow for pre-defined conditions to be simulated at scale, large-scale behavior to be observable, and resulting recommendations to be derived from the large-scale behavior.

5.1.Process of Randomized WBS

The randomized WBS script was developed in Visual Basic for Applications for Microsoft Excel. The process utilized in this script follows these steps:

1. Define the parameters of the simulation including the maximum number of WBS elements, the number of iterations per WBS size, and the statistics of the lowest-level WBS elements.
2. Randomly generate WBS's of varying size across multiple iterations.
3. Save the results of each iteration and calculate the resulting top-level CV of each WBS.
4. Examine the behavior of the top-level CV at scale.

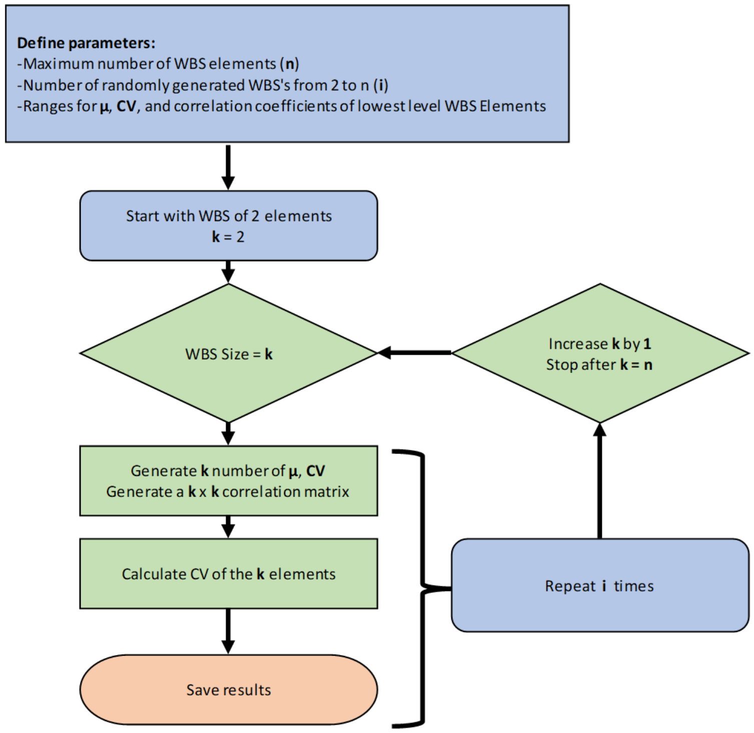


Figure 6 – Randomized WBS Process

Figure 6 above outlines the process used in the script. WBS size ranged from two to sixty lowest level elements, with forty WBS's randomly generated for each WBS size. While not shown in this technical paper, increasing the number of iterations or maximum WBS size did not yield different results significant to this study.

5.2. Sampling Statistics of Subordinate WBS Elements

An important step in generating these randomized WBS's, is to define parameters for sampling random values for the statistics of the lowest-level WBS elements. These parameters will define the behavior of the simulated cost estimators building a cost estimate. The first set of parameters to define are the mean values for each lowest level WBS elements. The mean values for the lowest level WBS elements are randomly sampled from a uniform distribution that enables a top-level sum on the scale of a Department of Defense Acquisition Category I program ($3.6B-$5.8B sum of Development and Procurement costs). The choice of this top-level sum is ultimately irrelevant for this study since preliminary simulations did not result in differing behavior and since CV is a statistic that scales with the mean value. For a WBS with n elements, the mean values for each lowest level WBS element are sampled from a uniform distribution with a lower bound of $\$3.6B/n$ and an upper bound of $\$5.8B/n$.

The more meaningful statistics to define parameters for are the CVs for each lowest level WBS element and the values for the correlation matrix of each WBS. For the CV of each lowest level WBS element, an exponential function was used to scale down the mean CV as the number of WBS elements increased. This was implemented because as the WBS is broken down into smaller elements, the scope of each element decreases and the level of certainty for that element increases. The scope of a cost element decreases as the fidelity and size of the WBS increases without changing the total scope of the overarching estimate, resulting in lower cost spreads for the subordinate cost elements of the WBS. For a WBS with n lowest level elements, the CV for each lowest level element is sampled from a normal distribution with a mean of 0.75 $e^{-0.025*(n-2)}$ and a standard deviation of 10% of the mean. Figure 7 depicts the sampling parameters for the CV of lowest level elements.

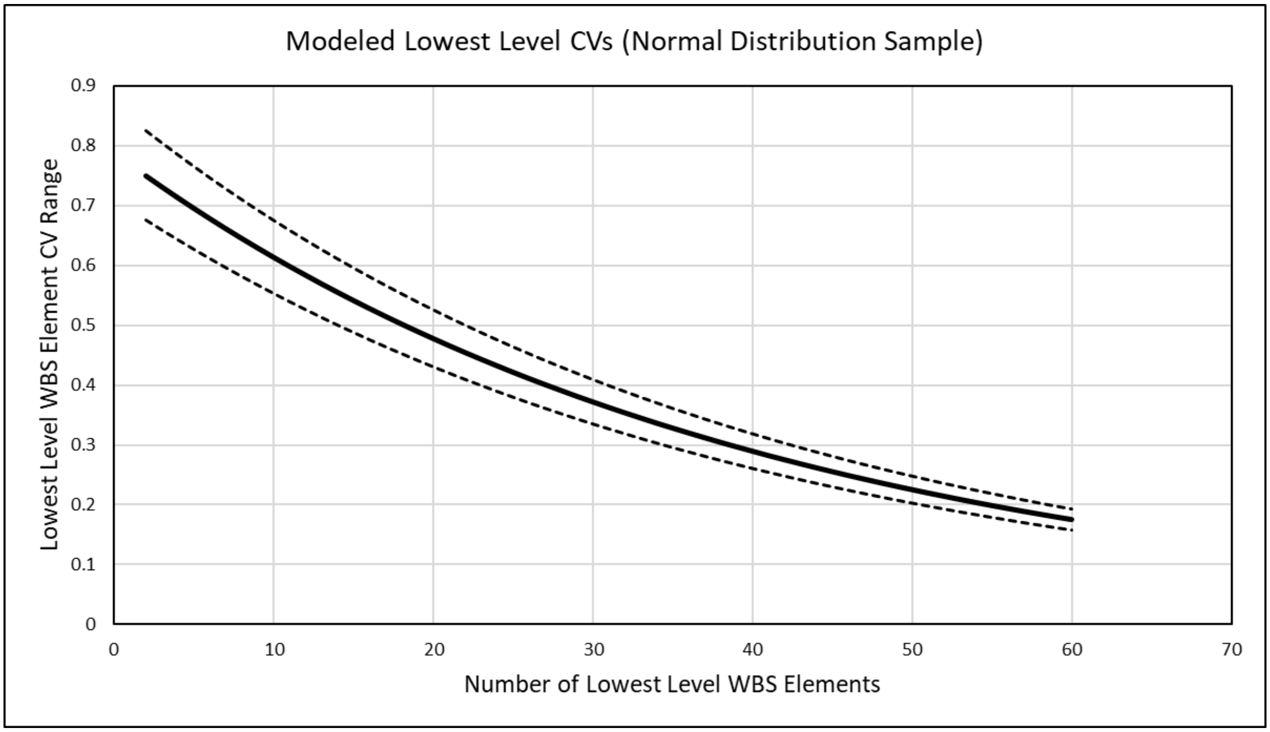


Figure 7 – Sampling Parameters of Lowest Level WBS Element CV

The last parameter to define is the level of correlation between lowest level WBS elements. For this parameter, the United States Air Force (USAF) Cost and Risk Analysis Handbook was utilized as it provides a rule of thumb for assigning correlation coefficients when the underlying correlation between elements is unknown. The USAF Cost and Risk Analysis Handbook recommends the following [4, p. 24]:

- 2-4 elements: correlation coefficients range from 0.5 to 0.75
- 5-19 elements: correlation coefficients range from 0.25-0.5
- 20 or more elements: correlation coefficients range from 0.1-0.25

This step-down range is used for defining the parameters of the sampled correlation coefficients. For a WBS of size n the correlation coefficients r are sampled from the following uniform distributions:

$$r \sim \begin{cases} Uniform(0.5, 0.75), & 2 \leq n \leq 4 \\ Uniform(0.25, 0.5), & 5 \leq n \leq 19 \\ Uniform(0.1, 0.25), & n \geq 20 \end{cases} \qquad (8)$$

The sampling pattern of the correlation coefficients in equation (8) are shown below in Figure 8.

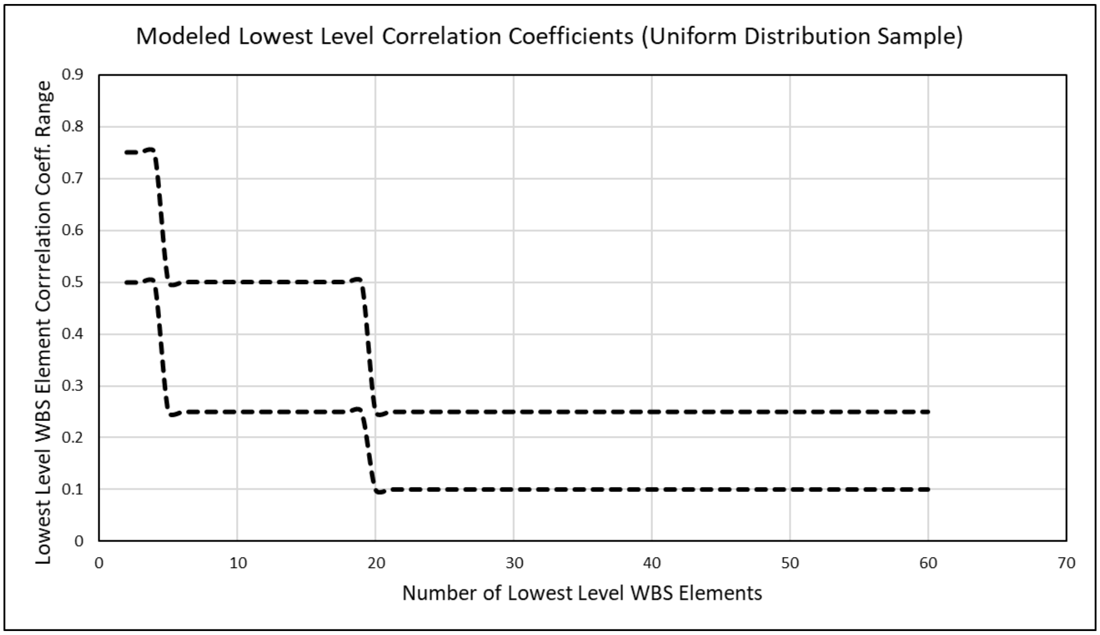


Figure 8 – Sampling Parameters of Lowest Level WBS Element Correlation Coefficients

5.3. Large-Scale Results of Simulations

Running large-scale simulations of randomized WBS's under these conditions yielded results which aligned with the large WBS perturbation, as the average top-level CV of each WBS trended downwards as the size of the WBS increased. Figure 9 is a graph of the average top-level CV by WBS size being plotted against the size of the WBS.

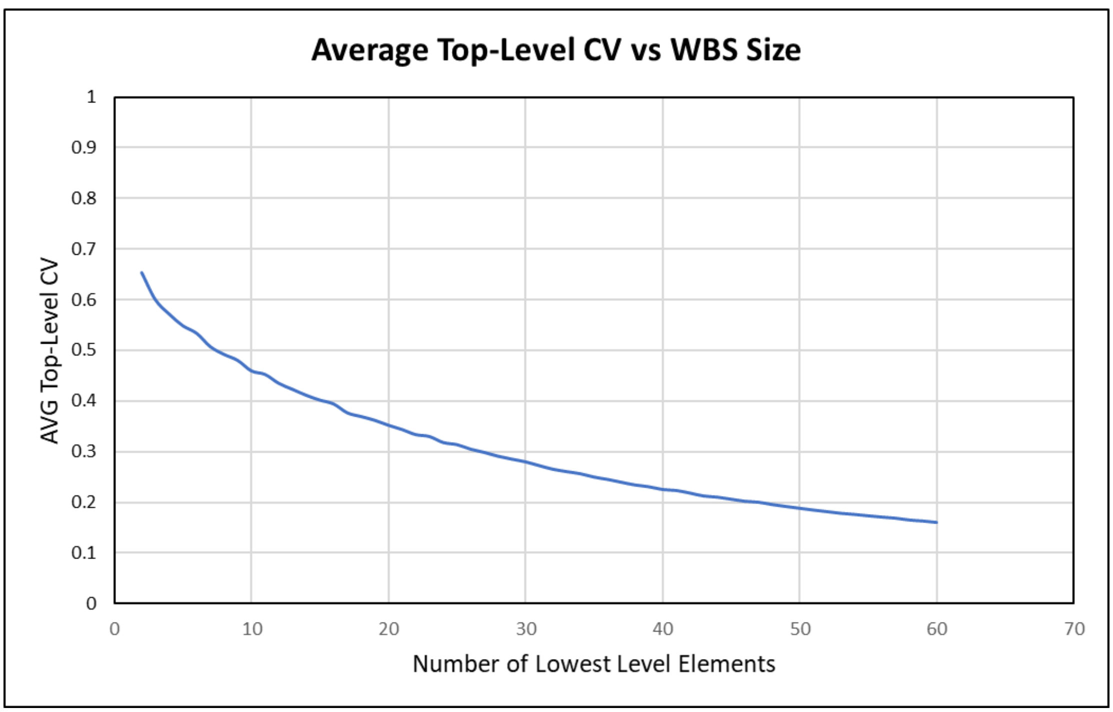


Figure 9 – Average Top-Level CV vs WBS Size

It is important to note that the downward trend shown in Figure 9 is partly a consequence of the parameters for lowest level WBS element CVs depicted in Figure 7. A less aggressive ramp down of lowest level element CV would result in a less aggressive downward trend of the average top-level CV, as indicated by the high CV perturbation, however it would be unrealistic to observe a risk-adjusted cost model with many lowest level WBS elements each with a CV greater than 0.35.

Additional analysis of this large-scale simulation includes an examination of the average ratio between top-level CV and the weighted average CV of the lowest level elements in that iteration, using the mean value of the lowest-level elements as the weighting factor. One method to determine if correlation has been underestimated is to examine the ratio between the top-level CV and the weighted average. The top-level CV will always be less than or equal to the weighted average CV of the lowest level elements, but a significant deviation represents underestimated cost uncertainty and is likely due to not having a sufficient level of correlation between input variables. [5]

The plot in Figure 10 shows the average ratio of top-level CV to the weighted average CV of the lowest level elements by WBS size, as well as dashed lines for +/− 25% adders from the average line. Cost estimators should use this plot to determine if the resulting top-level CV is too high or too low based on the size of the WBS. Ratios exceeding the dashed bounds of the plot indicate under correlated cost input variables if the ratio is below the lower bound or indicate over correlated cost input variables if the ratio is above the upper bound for a given WBS size.

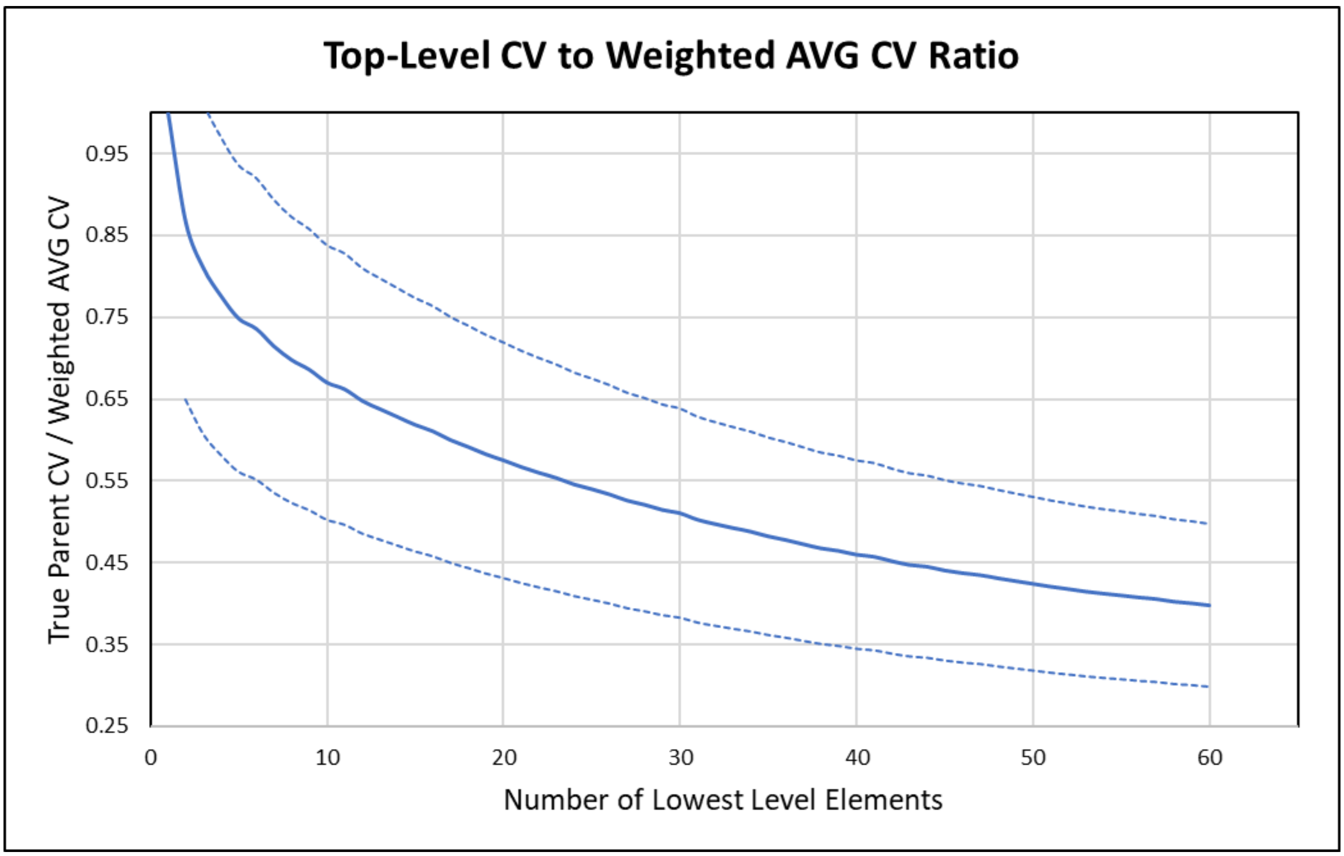


Figure 10 – Average Ratio of Top-Level CV to Weighted Average CV vs WBS Size

AuGUR CONSULTING

5.4. WBS Size Rule of Thumb

The top-level CV behavior at-scale can be paired with typical CV ranges across acquisition lifecycle to provide a recommendation on the size of the WBS which aligns to acquisition phase and the qualitative level of certainty on project requirements. A USAF Institute of Technology paper reviewed cost growth factors from actual DoD Programs of Record to determine appropriate ranges for top-level CV across program lifecycle [4, p. 80]. Cost growth factors represent deviations from actual program costs against the estimated cost at each acquisition phase milestone, which is ultimately an aspect of project management that a cost model's CV is intended to address.

USAF IT Research Paper	
Estimate Type	**CV Range**
Milestone A	0.41 - 0.74
Milestone B	0.31 - 0.54
Milestone C	0.23 - 0.32

Table 8 – USAF Institute of Technology CV Ranges by Acquisition Lifecycle

Using the appropriate ranges for CV across lifecycle shown in Table 8, along with the behavior of CV at scale, the following recommendation can be made for WBS size across acquisition lifecycle. This is also aligned with a qualitative level of overall project requirement certainty for use beyond the Department of Defense. For a bottom-up risk-adjusted cost model supporting a project with a high level of uncertainty, 2-14 lowest level WBS elements should comprise the WBS of the estimate. For cost estimates under the same conditions for a project with a medium level of uncertainty, 6-25 lowest level WBS elements is recommended. Lastly, an estimate for a project with a moderate or low level of uncertainty should have at least 24 lowest level WBS elements.

Recommended WBS Ranges	
Acquisition Phase	**Rec. WBS Size**
Milestone A/High Uncertainty	2 - 14 Lowest Level Elements
Milestone B/Medium Uncertainty	6 - 25 Lowest Level Elements
Milestone C/Modest Uncertainty	24+ Lowest Level Elements

Table 9 – Recommended WBS Size by Acquisition Phase/Project Requirements Certainty

Cost estimates are living documents that must be updated and maintained to provide continued meaningful insights to project managers and decision-makers. Having a WBS which reflects the overall level of certainty for a project is critical to ensuring that cost uncertainty is not underestimated.

5.5. Implications for Project Management

Managers should push for a level of cost variability in their delivered cost estimates that aligns with the maturity of a project. Early-stage projects should have a high degree of cost uncertainty which decays exponentially over time as the project matures and overall program certainty increases.

Cost estimates that accurately capture cost uncertainty can be used to realistically bound cost growth over project lifecycle, as well as provide more meaningful insights to derived values of management reserve and contingency. Managers should also ensure that the fidelity of a cost estimate accurately reflects the level of understanding of a project's requirements. Highly detailed cost estimates during the early-planning period of a project are likely to present a false sense of cost certainty and can trigger false flags if realistic cost growth exceeds underestimated cost growth bounds around an estimate.

6. Conclusion

Cost models should account for risk and statistical uncertainty in their outputs to appropriately forecast realistic cost growth or to derive realistic values for management reserve and contingency. The top-level CV of a cost model is used to normalize the variability of a cost model to allow for comparisons to other cost estimates and determine if cost uncertainty is appropriately captured in a model. Risk-adjusted cost models which follow the bottom-up method to applying uncertainty are potentially at risk of underestimating cost uncertainty. Using a derived equation for top-level CV, an understanding is gained for the underlying mechanisms in the model that control the top-level CV. Cost uncertainty of the lowest-level WBS elements, the correlation between input variables, and overall structure and detail of the WBS are the model attributes that an estimator can examine when diagnosing a cost model with lower-than-expected cost variability. Using randomly generated WBS's, large scale behavior of top-level CV indicates that total or near-total cost variability decreases with the size of the WBS. Comparing large-scale behavior of top-level CV to the weighted average CV of lowest level elements provides triggers that indicate if sufficient correlation has been applied to a cost model. Lastly, the at-scale behavior of cost variability is used to provide recommendations to WBS size that aligns with observed cost growth factors in an effort to persuade cost estimators to develop cost models with a level of fidelity that aligns to overall project certainty.

References

[1] U.S. Government Accountability Office, GAO-20-195G Cost Estimating and Assessment Guide, Washington, DC: U.S. Government Accountability Office, 2020.

[2] Defense Acquisition University, "Work Breakdown Structure (WBS)," Defense Acquisition University, Fort Belvoir, VA, Latest Revision.

[3] P. R. Garvey, Probability Methods for Cost Uncertainty Analysis, Bedford, MA: Marcel Dekker, Inc., 2000.

[4] U.S. Air Force, "Cost Risk and Uncertainty Analysis Handbook," U.S. Air Force, Hanscom Air Force Base, MA, 2007.

[5] D. J. Andelin, "Diagnosing the Top Level Coefficient of Variation: An Alternative Approach," International Cost Estimating & Analysis Association, Annadale, VA, 2012.

[6] C. U. Shaun T. Carney, "Investigation into Risk and Uncertainty: Identifying Coefficient of Variation Benchmarks," Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio, 2013.

Appendix A: Monte-Carlo Based Cost Modeling

Monte-Carlo simulations are often used to calculate the risk and uncertainty of a cost model, as well as other business and financial decisions. Monte-Carlo simulations in cost applications rely on repeated random sampling of values from probability distributions in a model. Each random sampling is referred to as an iteration of the simulation. Sampled values within a single iteration are then run through the model to calculate outputs from that iteration, the results are saved, and then new values are sampled to begin another iteration. The saved results across a finite number of iterations are finally aggregated to form an approximation of the output level probability distributions.

Figure A-1 is a graphic depicting the results of a Monte-Carlo simulation against the true results of the calculation. Within this scenario, two independent random variables adhering to normal distributions are being summed together to form a third, resulting random variable. The first probability distribution, shown in blue, has a mean of one and a standard deviation of 0.25. The second probability distribution, shown in orange, has a mean of three and a standard deviation of 0.1. The sum of those two random variables can be computed by hand without the use of any computational methods. The upper left-hand graph of Figure A-1 shows the probability density functions of these three distributions exactly.

The sum of the two random variables can also be modeled using a Monte-Carlo simulation as well. The graph in the upper right-hand corner of Figure A-1 has an overlayed histogram which represent random samples from the first iteration of a Monte-Carlo simulation. A random value is sampled from the first normal distribution, shown in blue, and another random value is sampled from the second normal distribution, shown in orange. These two random values are then summed together, a result which is shown by the black bar. This result is saved, and then four more iterations are run. The results of the first five iterations are shown in the bottom left-hand corner of Figure A-1. By increasing the number of iterations to 30, shown in the bottom left-hand corner of Figure A-1, the frequency plots begin to converge to the true shape of the probability density functions. Increasing the number of iterations will result in a convergence to the distribution of the sum of the two random variables. This is a simple calculation of only two independent random variables that is relatively easy to compute exactly by-hand, but larger and more complex models are more difficult and computationally taxing to compute exactly. Monte-Carlo simulations for these larger and more complex models follow a similar process to approximate resulting probability distributions of the model, in a computationally efficient manner.

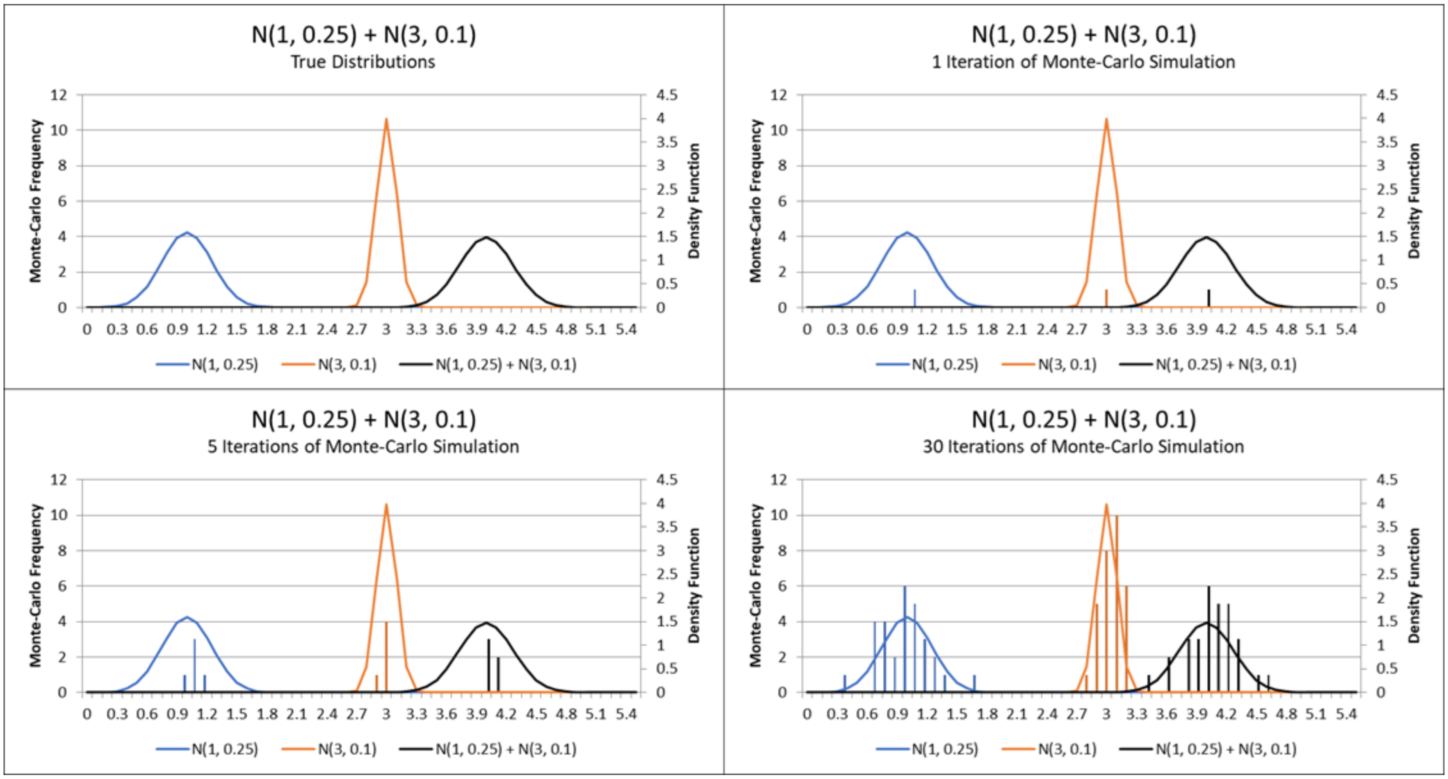


Figure A-1 – Monte-Carlo Sampling Illustrative Example

For a bottom-up risk adjusted cost model, a single iteration will randomly sample a value for each input variable of the model, calculate the cost output using those randomly sampled values, and then save the result of that iteration. Compiling the results of multiple iterations will result in approximated probability distributions at each level of the WBS. Cost estimators then use these probability distributions to generate cost figures at specific risk-levels that are appropriate for the use-case of the estimate. These cost figures represent snapshots of the output-level probability distribution at a specific confidence level.

Appendix B: Special Case of Independence

A potential issue in bottom-up risk-adjusted cost models, is that correlation between input variables is underrepresented or not accounted for at all. This absence of sufficient correlation will translate upwards to the WBS and converge to a model where WBS cost elements are independent of one another. To understand how under correlating variables in a bottom-up risk-adjusted cost model impacts the cost uncertainty at the total or near-total level, an examination of independence at the WBS level is required.

For lowest level WBS elements, X_i, in a cost model summing to the top-level, Z, assume each X_i are independent of each other. This is represented in the CV equation by the values of $r_{i,j}$ in the correlation matrix. When each X_i are independent, $r_{i,j}$ is equal to zero when $i \neq j$ and one when $i = j$. Plugging in these values of the Pearson correlation coefficients into equation (7) results in:

$$CV_{Z,\ independent} = \frac{\sqrt{\sum_{i=1}^{n}(CV_{X_i}\,\mu_{X_i})^2}}{\sum_{i=1}^{n}\mu_{X_i}} \tag{B-1}$$

For most applications of cost modeling, the inclusion of independence among WBS level probability distributions will result in a lower CV than if correlation is applied to input variables of the model. To demonstrate this, the following statement will be evaluated to determine the specific scenario where independence of WBS elements results in a wider spread than the inclusion of correlation:

$$CV_{Z,\ independent} > CV_Z \tag{B-2}$$

$$\frac{\sqrt{\sum_{i=1}^{n}(CV_{X_i}\,\mu_{X_i})^2}}{\sum_{i=1}^{n}\mu_{X_i}} > \frac{\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} r_{i,j}\,(CV_{X_i}\,\mu_{X_i})\,(CV_{X_j}\,\mu_{X_j})}}{\sum_{i=1}^{n}\mu_{X_i}} \tag{B-3}$$

$$\sum_{i=1}^{n}(CV_{X_i}\,\mu_{X_i})^2 > \sum_{i=1}^{n}\sum_{j=1}^{n} r_{i,j}\,(CV_{X_i}\,\mu_{X_i})\,(CV_{X_j}\,\mu_{X_j}) \tag{B-4}$$

The sum on the right-hand side of equation (B-4) can be expanded and divided into parts of the sum where $i = j$ and parts where $i \neq j$. When $i = j$, the sum of the equation reduces to the independent case plus the impact of correlated elements in the WBS. Let E be the set of elements in the sum on the right-hand side where $i \neq j$, or more exactly let the set be defined as: $E = \{ (i,j) \mid i \neq j; 1 \leq i,j \leq n \}$. Equation (B-4) becomes:

$$\sum_{i=1}^{n} (CV_{X_i}\, \mu_{X_i})^2 > \sum_{i=1}^{n} (CV_{X_i}\, \mu_{X_i})^2 + \sum_{(i,j) \in E} r_{i,j}\, (CV_{X_i}\, \mu_{X_i})\, (CV_{X_j}\, \mu_{X_j}) \qquad \text{(B-5)}$$

$$0 > \sum_{(i,j) \in E} r_{i,j}\, (CV_{X_i}\, \mu_{X_i})\, (CV_{X_j}\, \mu_{X_j}) \qquad \text{(B-68)}$$

Since each $\left(CV_{X_i}\mu_{X_i}\right)$ in equation (B-68) is non-negative in cost modeling applications, the above relationship implies a predominance of inverse/negative correlation at the WBS level of the cost model which is an unlikely scenario in practice. The inclusion of correlation in a risk-adjusted cost model will increase the variability of cost output at the total or near-total level, unless there is a majority of inverse correlation in the cost model which is a rare occurrence.

Appendix C: Perturbations to Baseline Case

Increased CV of WBS Element

The first perturbation to examine is the case when one of the elements in the WBS has a higher level of variability. This is modeled by increasing the CV, and the standard deviation by proxy, of one of the lowest-level WBS elements. The CV equation indicates that since the CV is in the numerator, increasing the CV of a WBS element will increase the overall spread at the total level. This is an intuitive result, and the level of increase can be shown by doubling the value of the standard deviation for one WBS element.

Baseline WBS			
WBS	Mean	Stan. Dev.	CV
Z	300		
X1	100	10	0.1
X2	100	10	0.1
X3	100	10	0.1

Increased CV of WBS Element			
WBS	Mean	Stan. Dev.	CV
Z	300		
X1	100	20	0.2
X2	100	10	0.1
X3	100	10	0.1

CV Calculation	
Baseline	0.070711
Increased CV of WBS Element	0.097183
% Δ CV	37%

Correlation Matrix			
	X1	X2	X3
X1	1	0.25	0.25
X2	0.25	1	0.25
X3	0.25	0.25	1

Correlation Matrix			
	X1	X2	X3
X1	1	0.25	0.25
X2	0.25	1	0.25
X3	0.25	0.25	1

Table C-1 – Increased CV of WBS Element Perturbation

By doubling the CV of X_1 from 0.1 to 0.2, the CV of Z increased from 0.0707 to 0.0972, which is approximately a 37% increase in the CV of Z from the baseline case.

Large Mean

Increasing the mean of a single WBS element is the next perturbation to examine. For this scenario, the mean of X_1 will be doubled from 100 to 200. Doubling the mean of a single element will have the immediate impact of increasing the standard deviation of X_1 to keep the CV of X_1 constant, as well as increasing the mean value of the top-level WBS element Z. Unlike increasing the CV of a single WBS element, increasing the mean of a single WBS element does not have an obvious result when reviewing the equation. The numerator of the CV equation will increase in value, but this increase is bounded by an increase in the denominator as well.

Baseline WBS			
WBS	Mean	Stan. Dev.	CV
Z	300		
X1	100	10	0.1
X2	100	10	0.1
X3	100	10	0.1

Large Mean			
WBS	Mean	Stan. Dev.	CV
Z	400		
X1	200	20	0.1
X2	100	10	0.1
X3	100	10	0.1

CV Calculation	
Baseline	0.070711
Large Mean	0.072887
% Δ CV	3%

Correlation Matrix			
	X1	X2	X3
X1	1	0.25	0.25
X2	0.25	1	0.25
X3	0.25	0.25	1

Correlation Matrix			
	X1	X2	X3
X1	1	0.25	0.25
X2	0.25	1	0.25
X3	0.25	0.25	1

Table C-2 – Large Mean Perturbation

The result of this perturbation is an increase in the CV of Z from 0.0707 to 0.0729, which is an increase of approximately 3% from the baseline case. This is a minimal increase to the spread at the top-level of the WBS, given that the perturbation is significant to mean of the top-level sum.

Large Mean (Normalized)

An issue with the previous perturbation is that the mean value of the top-level element Z increased significantly, which may not be a logical alteration to a cost model when remedying underrepresented cost variability in a model. To combat this potentially confounding aspect of the previous perturbation, another scenario is established which normalizes the mean values of the other two elements of the WBS so that the mean value of Z is retained. The mean of X_1 is again doubled to 200, and the means of X_2 and X_3 are equally decreased to 50 so that the mean value of Z remains at 300. Note that the values of the standard deviation for each lowest level WBS element is adjusted to preserve the baseline CV of each element.

Baseline WBS			
WBS	Mean	Stan. Dev.	CV
Z	300		
X1	100	10	0.1
X2	100	10	0.1
X3	100	10	0.1

Large Mean (Normalized)			
WBS	Mean	Stan. Dev.	CV
Z	300		
X1	200	20	0.1
X2	50	5	0.1
X3	50	5	0.1

CV Calculation	
Baseline	0.070711
Large Mean (Normalized)	0.079057
% Δ CV	12%

Correlation Matrix			
	X1	X2	X3
X1	1	0.25	0.25
X2	0.25	1	0.25
X3	0.25	0.25	1

Correlation Matrix			
	X1	X2	X3
X1	1	0.25	0.25
X2	0.25	1	0.25
X3	0.25	0.25	1

Table C-3 – Large Mean (Normalized) Perturbation

By increasing the mean of X_1 and decreasing the means of X_2 and X_3 to preserve the mean of Z increases the CV of Z to 0.0791, which is approximately a 12% increase from the baseline. The resulting CV of this perturbation is substantially higher than the CV of the previous perturbation, but still less impactful than changing the CV of a single WBS element. An interpretation of this scenario is that elements in a WBS are grouped in a matter that preserves their sum but increases the level of overall variability in the WBS.

Large WBS

This perturbation demonstrates the impact of increasing the number of elements in a WBS, while preserving the sum at the top-level element and maintaining the CV of each lowest level WBS element. This can be thought of as increasing the fidelity of the estimate by furthering decomposing the existing WBS elements into smaller components. For this perturbation, the WBS is altered to the sum of six lowest level elements, all with mean values of 50 so that their sum matches the sum in the baseline case. The equation for CV does not immediately provide any insight to the expected results of this perturbation.

Baseline WBS			
WBS	Mean	Stan. Dev.	CV
Z	300		
X1	100	10	0.1
X2	100	10	0.1
X3	100	10	0.1

Large WBS			
WBS	Mean	Stan. Dev.	CV
Z	300		
X1	50	5	0.1
X2	50	5	0.1
X3	50	5	0.1
X4	50	5	0.1
X5	50	5	0.1
X6	50	5	0.1

CV Calculation	
Baseline	0.070711
Large WBS	0.061237
% Δ CV	-13%

Correlation Matrix			
	X1	X2	X3
X1	1	0.25	0.25
X2	0.25	1	0.25
X3	0.25	0.25	1

Correlation Matrix				
	X1	X2	...	X6
X1	1	0.25	0.25	0.25
X2	0.25	1	0.25	0.25
...	0.25	0.25	1	0.25
X6	0.25	0.25	0.25	1

Table C-4 –Large WBS Perturbation

Increasing the size of the WBS while maintaining the top-level sum results in a lower top-level CV, as the CV dropped approximately 13% down to 0.0612 from the baseline case. This result indicates that a larger WBS with a high volume of small lowest level elements results in a lower level of dispersion at the total or near-total level compared to a model with a WBS with the same mean value but less numerous lowest level WBS elements. Another interpretation of this result is that increasing the level of fidelity in a WBS is an opposite mechanism to the previous perturbation where stronger grouping of lowest level elements in a WBS increases the overall spread of the model.

AuGUR CONSULTING

Strong Correlation

The last aspect of the CV equation to examine, is the impact of correlation on the modeled cost spread at the total or near-total level. For this perturbation, the only change is that the correlation coefficient between a single pair of lowest level WBS elements will be doubled. Since the correlation coefficients only appear in the numerator of the CV equation, the top-level CV will increase because of this perturbation.

Baseline WBS			
WBS	Mean	Stan. Dev.	CV
Z	300		
X1	100	10	0.1
X2	100	10	0.1
X3	100	10	0.1

Strong Correlation			
WBS	Mean	Stan. Dev.	CV
Z	300		
X1	100	10	0.1
X2	100	10	0.1
X3	100	10	0.1

CV Calculation	
Baseline	0.070711
Strong Corr	0.074536
% Δ CV	5%

Correlation Matrix			
	X1	X2	X3
X1	1	0.25	0.25
X2	0.25	1	0.25
X3	0.25	0.25	1

Correlation Matrix			
	X1	X2	X3
X1	1	0.25	0.5
X2	0.25	1	0.25
X3	0.5	0.25	1

Table C-5 –Strong Correlation Perturbation

As expected, the top-level CV increased to 0.0745 which represents a 5% increase from the CV of the baseline case. While this is a minor overall impact to the top-level spread, it is more important to note that increasing a single correlation coefficient is a relatively insignificant perturbation that causes a disproportionate change to the top-level spread of the cost model.

Independence (No Correlation)

The derived equation for top-level CV and extrapolating the results of the prior perturbation both indicate that decreasing the correlation between lowest level WBS elements in a cost model will result in a lower overall CV at the total or near-total level. This perturbation is designed to illustrate the impact of no correlation in a risk-adjusted cost model. Independent lowest level WBS elements will have off-diagonal entries in the correlation matrix of zero, while the diagonal entries are one by default.

AuGUR CONSULTING

Baseline WBS			
WBS	Mean	Stan. Dev.	CV
Z	**300**		
X1	100	10	0.1
X2	100	10	0.1
X3	100	10	0.1

No Correlation			
WBS	Mean	Stan. Dev.	CV
Z	**300**		
X1	100	10	0.1
X2	100	10	0.1
X3	100	10	0.1

CV Calculation	
Baseline	0.070711
No Corr	0.057735
% Δ CV	**-18%**

Correlation Matrix			
	X1	**X2**	**X3**
X1	1	0.25	0.25
X2	0.25	1	0.25
X3	0.25	0.25	1

Correlation Matrix			
	X1	**X2**	**X3**
X1	1	0	0
X2	0	1	0
X3	0	0	1

Table C-6 –Independence (No Correlation) Perturbation

Independence of lowest level elements in a WBS results in a resulting top-level CV of 0.0577, or a 18% decrease in the top-level CV from the baseline case. Failure to apply correlation to a cost model can have a dramatic effect on the spread at the top-level of the WBS. This impact can be compounded when the cost model has many lowest level WBS elements.