UNCLASSIFIED

Elements of Cost Risk Analysis

A SYSTEMS ENGINEERING PERSPECTIVE

Paul R. Garvey, Ph.D.

January 2014

Department No.: J810
Project No.: 0714ECSE-PA

Document No: MP130729
Location: Bedford, Massachusetts

Corresponding author, pgarvey@mitre.org

UNCLASSIFIED

MITRE

ELEMENTS OF COST RISK ANALYSIS

SUMMARY

Paul R. Garvey[1]

Cost is a driving consideration in decisions that determine how systems are developed, produced, and sustained. Understanding how risk affects a system's cost is critical to these decisions. The process of identifying, measuring, and interpreting these effects is known as *cost risk analysis*.

This paper was written for a US government agency to provide the foundations, objectives, and benefits of cost risk analysis in the acquisition and engineering of systems. Major elements of the analysis are described and modern practices are presented. For the experienced cost analyst, this material may serve as a refresher. For those new to the field, this material provides key methods and practice points essential for conducting, evaluating, or conveying a cost risk analysis[2].

The paper begins with an introduction to elementary concepts and key terms. This includes a discussion on the scope of cost risk analysis (what is captured, what is not captured), what it means to present and interpret cost as a probability distribution, and the insights cost risk analysis brings to decision-makers.

The main body of this paper presents two worlds within which cost risk analysis methods fall. They are Monte Carlo simulation methods and method of moments approaches. Introductions to these methods and approaches are provided. Popular tools that execute these techniques are summarized and nuances on their application to cost risk analyses are explained.

This paper clarifies issues associated with certain technical topics that arise in cost risk analysis. These include the use of normal or lognormal distributions to derive confidence intervals around point estimates, as well as how best to deal with correlation between cost elements of a program. Common mistakes, pitfalls, and guidance on conducting a cost risk analysis are discussed, as well as conveying its findings.

The paper concludes with summary practice points and recent research by the Naval Center for Cost Analysis that produced historical cost risk measures from numerous major defense programs[3]. These histories provide new insights from data on past programs that can guide and shape cost risk analyses.

---

[1] Corresponding author address: pgarvey@mitre.org

[2] For a comprehensive mathematical treatment of cost risk analysis, refer to the text *Probability Methods for Cost Uncertainty Analysis: A Systems Engineering Perspective*, P. R. Garvey, Chapman-Hall/CRC Press, 2000 (2nd edition in-progress).

[3] The collection and analysis of historical data for use in cost risk analysis was under the auspices of Ms. Wendy Kunc, (Deputy Assistant Secretary for Cost and Economics, Office of the Assistant Secretary of the Navy (Financial Management & Comptroller) and Executive Director of the Naval Center for Cost Analysis (NCCA). The reader is also directed to https://www.ncca.navy.mil/tools/tools.cfm for further information about the NCCA historical cost risk data.

ELEMENTS OF COST RISK ANALYSIS

A SYSTEMS ENGINEERING PERSPECTIVE[4]

"It's not what you don't know that hurts you – it's what you DO know that isn't true." - Dr. Stephen A. Book (1995)

1.0 PURPOSE

This paper was written for a US government agency to provide the foundations, objectives, and benefits of cost risk analysis in the acquisition and engineering of systems. Major elements of the analysis are described and modern practices are presented. For the experienced cost analyst, this material may serve as a refresher. For those new to the field, this material provides key methods and practice points essential for conducting, evaluating, or conveying a cost risk analysis[5].

1.1 INTRODUCTION

Decreasing budgets, changing requirements, rapidly evolving technologies, and heightened competition for resources are pressures that bring persistent attention to program cost. Given this, cost analysis is a critical activity in engineering today's systems. It provides the economic perspectives necessary to acquire, deploy, and maintain systems over their life cycles.

In systems engineering, costs are estimated to reveal the economic significance of technical and programmatic choices that guide procuring a system that is affordable, cost-effective, and risk managed. Identifying risks enables decision-makers to develop, execute, and monitor management actions based on the knowledge of potential cost consequences of inactions. Together, cost and cost risk analysis are undertaken to address paramount considerations of affordability, cost-effectiveness, and risk. Affordability addresses the question: "*Can the system be procured with the funds available?*" Cost-effectiveness addresses the question: "*Does the system represent the best use of funds?*" Risk addresses the question: "*What is the chance the planned or budgeted cost of the system will be exceeded?*" Given this, the purpose of cost risk analysis is to (1) enable the early and continuous identification of cost risk driving elements and (2) produce a defensible assessment of the level of cost to plan or budget, so there is reasonable confidence in assuring program affordability and cost-effectiveness.

Risk analysis is a complex and inseparable part of cost analysis. Many different elements of a program's technical baseline and cost estimate are involved. This includes technology maturity, supply chain integrity, quantities, schedules, and acquisition considerations. The mathematics of risk analysis can be advanced – often utilizing concepts of correlation, probability distributions, and means and variances. Conveying risk analysis findings clearly and concisely to audiences with broad backgrounds is a challenging yet crucial aspect of the process. It is critically important to step back from the analysis to understand what the findings really mean, whether risks have been adequately captured, and what can be done to reduce their potential negative consequences.

Defense Costs

A recent GAO[6] assessment of the 86 programs that make up the 2012 portfolio of major defense acquisition programs found a net cost decrease over the past year, nearly all of which was attributable to quantity reductions stemming from program cancelations and restructuring[7]. GAO further observed that

[4] Paul R. Garvey, Ph.D., Chief Scientist, Center for Acquisition and Systems Analysis, MITRE Corporation, pgarvey@mitre.org.
[5] For a comprehensive mathematical treatment of cost risk analysis, refer to the text *Probability Methods for Cost Uncertainty Analysis: A Systems Engineering Perspective*, P. R. Garvey, Chapman-Hall/CRC Press, 2000 (2nd edition in-progress).
[6] GAO-13-294SP: Defense Acquisitions, Assessments of Selected Weapon Programs, March 2013.
[7] Excerpted from GAO-13-294SP, p. 1, March 2013.

over the past year, a majority of programs in the portfolio experienced increased buying power; the percentage of programs meeting cost targets discussed by GAO, DOD, and the Office of Management and Budget increased; and nearly all of the 10 largest programs in the portfolio with program baselines reported reductions in total estimated cost[8]. However, shown in Table 1, when measured against first full estimates, the portfolio experienced total acquisition cost growth of 38 percent. In addition, 12 programs delayed the delivery of their initial capabilities resulting in an average increase of nearly 1-month across the portfolio. The average delay in delivering initial capability was 27 months when measured against first full estimates[9].

Table 11: Cost and Schedule Changes for Programs in DOD's 2012 Portfolio

Fiscal year 2013 dollars in billions		
	5 year comparison (2007 to 2012)	**Since first full estimate (first full est. to 2012)**
Change in total research and development cost	$27 billion 12 percent	$94 billion 49 percent
Change in total procurement cost	$130 billion 14 percent	$294 billion 35 percent
Change in total acquisition cost	$158 billion 13 percent	$403 billion 38 percent
Average change in delivering initial capabilities	12 months 17 percent	27 months 37 percent

Table 1. Cost and Schedule Growth: DOD 2012 Program Portfolio[10]

Mentioned earlier, technology maturity can be a significant risk and cost growth driver. GAO reported in the past that programs working to mature technologies after the start of development while concurrently attempting to mature a system's design and prepare for production are at higher risk of experiencing cost growth and schedule delays. GAO observed that those programs tend to have higher cost growth than programs that start system development with mature technologies. The GAO analysis indicates the average rate of development cost growth for those programs that started with immature technologies is 86 percent, while the average growth rate for development costs is about half that amount for programs that began with their critical technologies at least nearing maturity[11].

A recent initiative to address cost growth in defense acquisitions is the Weapon Systems Acquisition Reform Act (WSARA), signed into law in 2009[12]. One aim of WSARA is to improve cost realism by requiring acquisition programs to budget at a high degree of confidence, such that it can be completed without the need for significant cost adjustments at a later phase. Given this, WSARA requires programs measure their cost estimate confidence relative to the 80th percentile level. The methods and approaches to cost risk analysis described herein are key ways to meet this condition.

[8] Excerpted from GAO-13-294SP, p. 1, March 2013.
[9] Excerpted from GAO-13-294SP, p. 1, March 2013.
[10] Source: GAO-13-294SP , p. 169, March 2013. Data obtained from DOD's SARs and acquisition program baselines and directly from program offices. Not all programs had comparable cost and schedule data and these programs were excluded from the analysis where appropriate. Total acquisition cost includes research and development, procurement, acquisition operation and maintenance, and system-specific military construction costs. First full estimates are analogous to point estimate costs. Confidence intervals associated with these estimates are either not explicitly determined or not reported with SAR filings.
[11] Paragraph excerpted from GAO-13-294SP, p. 25, March 2013.
[12] Public Law 111-23, 2009. "Weapon Systems Acquisition Reform Act of 2009", 22 May 2009.

Cost Risk Analysis

Acquiring today's systems is more sophisticated and complex than ever before. Increasingly, systems are engineered by bringing together many separate systems, which, as a whole, provide a capability otherwise not possible. Systems are now richly connected. They involve and evolve webs of users, technologies, and systems-of-systems through environments that offer cross-boundary access to a wide variety of resources and information repositories. Today's systems create value by delivering capabilities over time that meet user needs for increased agility, robustness, and scalability. System architectures must be open to allow the insertion of innovation that advances the efficacies of capabilities and services to users.

Many systems no longer physically exist within well-defined boundaries. They are increasingly ubiquitous and operate as an enterprise of technologies and cooperating entities in a dynamic that can behave in unpredictable ways. Pervasive with these challenges are economic and budgetary realities that necessitate greater accuracy in the estimated life cycle costs and cost risks of acquiring these systems.

Systems engineering is more than developing and employing inventive technologies. Designs must be adaptable to change, flexible to meet user needs, and resource managed. They must be balanced with respect to performance and affordability goals while being continuously risk managed throughout a system's life cycle. Systems engineers and managers must also understand the social, political, and economic environments within which a system operates. These factors can significantly influence risk, affordability, design options, and investment decisions.

Applied early and continuously, risk analysis can expose events that, if realized, might impede an acquisition program from achieving its required cost, schedule, and performance goals. Risk analysis is more than identifying and quantifying the consequences of unwanted events on an acquisition program and its cost. It provides a context for bringing realism to technical and program decisions that shape a program's acquisition strategy and the cost-effectiveness of its long-term performance.

Why are there risks? Pressures to meet cost, schedule, and technical performance are the practical realities in acquiring systems. Illustrated in Figure 1, risk becomes an increasing threat when stakeholder expectations push what is technically or economically feasible. Thus, managing risk is managing the inherent contention that exists within and across these dimensions.

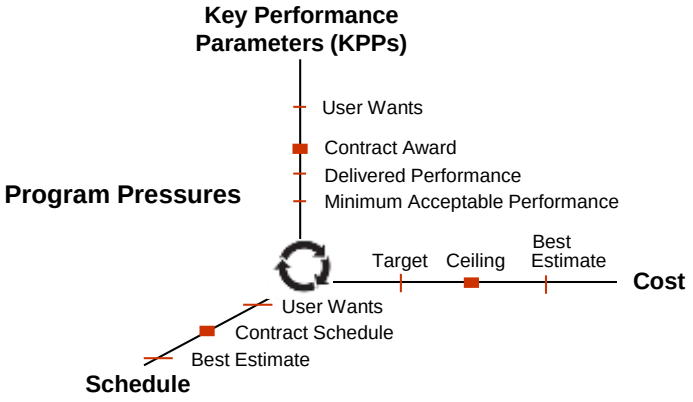


Figure 1. Pressures on a Program Manager's Decision Space

For cost risk analysis to shape and influence program decisions, it must provide insights into cost estimate confidence that are otherwise unseen. What does cost estimate confidence mean? In general, it is a statement of the sureness in an estimate along with a supporting rationale. The intent of cost risk analysis

is to enable statements on cost estimate confidence to be addressed, such as "*there is an 80 percent chance the program's actual cost will not exceed $250M*". How is cost estimate confidence measured?

Probability theory is an ideal formalism for deriving measures of cost estimate confidence. With it, a program's cost can be treated as an uncertain quantity – one sensitive to conditions and assumptions that change across its acquisition life cycle. Figure 2 illustrates the conceptual process for using probability theory to analyze cost uncertainty and produce measures of cost estimate confidence.

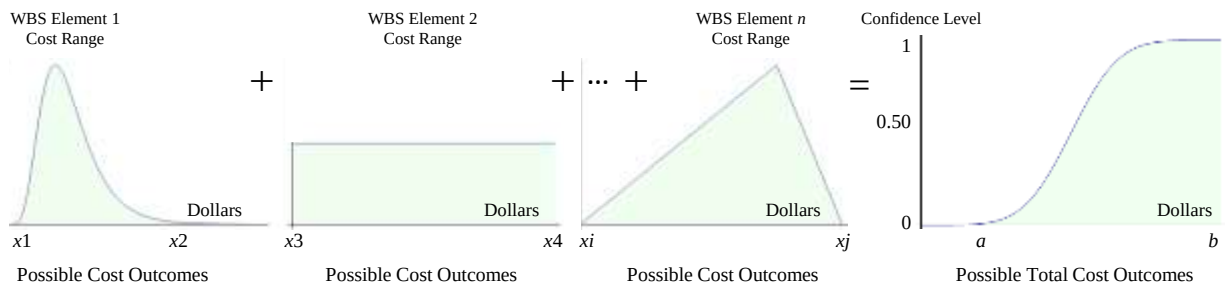


Figure 2. Cost Estimate Confidence: A Statistical Summation of Cost Uncertainty

In Figure 2, the uncertainty in the cost of each work breakdown structure[13] (WBS) element is expressed by a probability distribution to characterize its range of possible cost outcomes. These distributions are combined by probability methods to generate an overall distribution of the work breakdown structure's total cost, hereafter denoted by the notation $Cost_{WBS}$. This distribution is the range of total cost outcomes possible for the WBS, as it represents a program or system. How does the output from this analytical process enable confidence levels to be determined? Consider Figure 3.

Figure 3 illustrates a cumulative probability distribution of a work breakdown structure's total cost. It derives from an analysis like that in Figure 2. Cost estimate confidence is read from this distribution. For example, there is a 25 percent chance the program will cost less than or equal to $100M, a 50 percent chance the program will cost less than or equal to $151M, and an 80 percent chance the program will cost less than or equal to $214M. These are examples of statistical measures of cost estimate confidence.

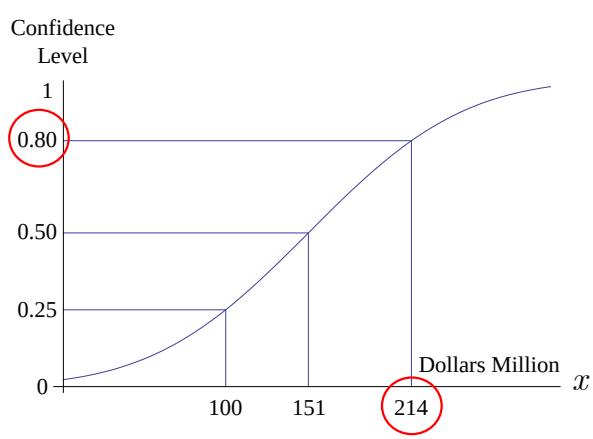


Figure 3. A Distribution of Cost Estimate Confidence

[13] Refer to military standard *MIL-STD-881C Work Breakdown Structures for Defense Material Items*, 3 October 2011, for information about work breakdown structures, how they are designed, and their roles in the systems engineering life cycle.

Figure 3 provides decision-makers an analytical basis for tradeoffs between a program's point estimate cost[14] and its confidence. For example, if a program's point estimate cost is $100M then Figure 3 reveals the amount of additional dollars needed to plan or budget the program at a desired or specified level of confidence. Clearly, the range of possible cost outcomes in Figure 3 is quite wide. Cost analysts can use such a finding to signal a review of the major cost risk drivers responsible for this variance, before settling too soon on a cost confidence level to plan or budget the program. Furthermore, the results presented in Figure 3 might spark a series of design options taken to reduce risk evident in the variance. Figure 4 shows how a cost risk analysis of a set of risk reducing design alternatives can be portrayed.

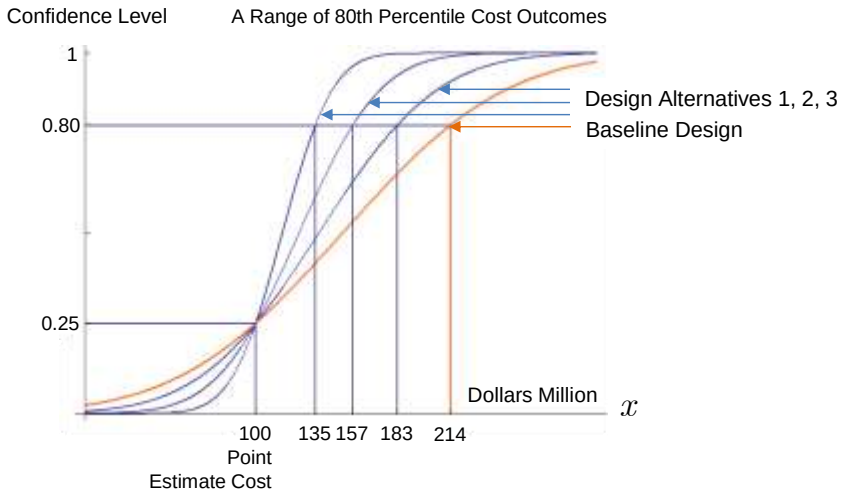


Figure 4. Reductions in Cost Risk from Competing Design Options

Discovering these findings and bringing them to decision-makers is a key outcome of cost risk analysis and how it best contributes to ensuring program affordability and cost-effectiveness. Expressing cost estimate confidence by a range of possible cost outcomes has high value to decision-makers. The extent of the cost range itself is a measure of cost uncertainty, which varies across the life cycle. One would expect uncertainty to be higher, and, therefore, the cost range to be wider, early in a program's life cycle. Identifying critical elements that drive a program's cost range offers opportunities for deploying risk mitigation actions in the early acquisition phases. In general, cost risk analysis enables the following:

Practice Point 1: *Establish a Cost and Schedule Risk Baseline* – Baseline probability distributions of program cost and schedule should be developed for a given system configuration (its technical baseline), acquisition strategy, and cost-schedule estimation approach. The baseline provides decision-makers visibility into potentially high-payoff areas for risk reduction initiatives. Baseline distributions assist in determining a program's cost and schedule that simultaneously have a specified probability of not being exceeded. They also provide decision-makers an assessment of the chance of achieving a budgeted (or proposed) cost and schedule, or cost for a given feasible schedule.

Practice Point 2: *Measure Cost Risk* – Cost risk analysis provides a basis for measuring the overall cost risk inherent to a program as a function of its specific uncertainties. This can be measured by the difference between the program point estimate cost and the cost at a predefined confidence level, as set by budgetary decisions or management policies.

[14] For purposes of this paper, the point estimate (PE) cost is the cost that does not include allowances for cost uncertainty. The PE cost is the sum of the WBS element costs summed across a program's work breakdown structure without adjustments for uncertainty. The PE cost is often developed from a program's cost analysis requirements description document (CARD).

Practice Point 3: *Conduct Risk Reduction Tradeoff Analyses* – Cost risk analyses can be conducted to study the payoff of implementing risk reduction initiatives on lessening a program's cost, schedule, and performance risks. Families of probability distribution functions, as shown in Figure 4, can be generated to compare the cost and cost risk impacts of competing design options or acquisition strategies.

Practice Point 4: *Document Program Risks and Risk Analysis Inputs* – The validity and influence of any cost risk analysis relies on the engineering and cost team's experience, judgment, and knowledge of their program's risks and uncertainties. Documenting the team's insights into these considerations is a critical part of the process. Without it, the veracity of the cost risk analysis is easily questioned. Details about the analysis methodology, especially assumptions, are important to document. The methodology must be technically sound, traceable, and offer value-added problem structure and insights otherwise not visible. Decisions that successfully reduce or eliminate risk ultimately rest on human judgment. This is aided by, not directed by, the methods in this paper.

2.0 FOUNDATIONS AND OBJECTIVES OF COST RISK ANALYSIS

This section presents foundations of cost risk analysis, its objectives, and key terms. The discussion includes the types of uncertainties captured in the analysis, reasons for viewing cost as a probability distribution, and the importance of measuring cost estimate confidence.

Cost estimates of future programs are inherently uncertain. Unstable requirements, ambiguity in a system's technical definition, and errors in cost estimation methods are among the factors that contribute to uncertainty. Recognizing this, a cost estimate is best accompanied by a range of possible cost outcomes expressed as a confidence interval. Range information conveys a sense of the completeness, stability, and maturity of a technical baseline. In addition, expressing a cost estimate as a range of possible cost outcomes enables management to take a position with respect to risk. For budgeting, the view may be risk-averse, i.e., reserve additional funds for overruns. For negotiating, the position may be to transfer ownership of cost risk to the group best able to address its mitigation. In conducting design trades to reduce risk, or achieve affordability goals, range information can identify whether cost uncertainties between choices wash away any cost differences between them.

When the cost of a program is estimated, decision-makers often ask, "*What is the chance the actual program cost will overrun the estimate?*" "*How much could it overrun?*" "*What are the risks and how do they drive cost?*" Cost uncertainty analysis provides decision-makers insight into these and related questions. Throughout a system's life cycle, cost uncertainty analysis provides motivation and structure for the vigorous management of risk. When appropriately communicated to decision-makers, the insights produced by the analysis direct management's attention to critical program risk-drivers. This enables risk mitigation strategies to be defined and implemented in a timely and cost-effective manner.

2.1 THREE TYPES OF UNCERTAINTY

Cost risk analysis had its genesis in a field known as military systems analysis [Hitch, 1955], founded in the 1950s at RAND Corporation. Shortly after World War II, military systems analysis evolved as a way to aid defense planners with long-range decisions on force structure, force composition, and future theaters of operation. Cost became a critical consideration in military systems analysis models and decision criteria. However, cost estimates of future military systems, particularly in the early planning phases, were often significantly lower than the actual cost or an estimate developed at a later phase. In the book *Cost Considerations in Systems Analysis* [Fisher, 1971] this difference is attributed to the presence of uncertainty; specifically, requirements uncertainty and cost estimation uncertainty.

Requirements uncertainty can originate from unforeseen changes in the system's mission objectives, in performance requirements necessary to meet mission objectives, or in the business or geopolitical landscapes that affect the need for the system. Requirements uncertainty often results in changes to the hardware-software configuration, which is sometimes called the system's architecture.

Cost estimation uncertainty can originate from inaccuracies or imprecision in cost-schedule estimation models, from the misuse (or misinterpretation) of cost-schedule data, or from misapplied cost-schedule estimation methodologies. Economic uncertainties that influence the cost of technology, the labor force, or geo-political policies further contribute to cost estimation uncertainty.

Uncertainty is also present in elements that define a system's technical baseline. This is referred to as *system definition uncertainty*. Examples include uncertainties in the size of the development software, the extent to which code from another system can be reused, the number of servers to procure, or the delivered weight of an end-item (e.g., a satellite).

These are the three types of uncertainties considered in cost risk analysis. Figure 5 illustrates how they are related. Shown in Figure 5A, requirements are defined that produce a specific system configuration; from this, cost risk analysis then captures uncertainties associated with the system definition (technical baseline) and cost estimation methods associated with that configuration. The *n* system configurations shown in Figure 5B are in response to requirements uncertainty. If changes in requirements result in a markedly different system configuration from which costs were originally estimated, then it is best to treat this as an alternative configuration – one necessitating a new cost estimate with an accompanying cost risk analysis. The new analysis must capture the system definition (technical baseline) and cost estimation uncertainties unique to the configuration being considered.

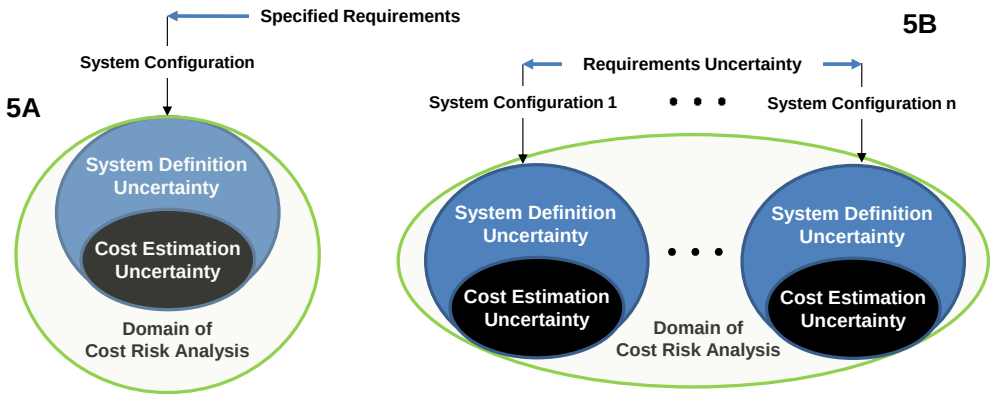


Figure 5. Uncertainties Addressed in Cost Risk Analysis [Garvey, 2000]

2.2 COST AS A PROBABILITY DISTRIBUTION

Cost estimates are highly sensitive to many conditions and assumptions that change frequently across a program's life cycle. Examining the change in cost subject to varying certain conditions, while holding others constant, is known as sensitivity analysis. Sensitivity analysis is an excellent a way to isolate cost drivers, however, it is a deterministic procedure defined by a postulated set of what-if scenarios. Sensitivity analysis alone does not offer decision-makers insight into the question "*What is the chance of exceeding a particular cost in the range of possible costs?*" A probability distribution is an ideal way to address this question. In the context of cost risk analysis, a probability distribution is a mathematical rule associating a probability to each cost in a range of possible cost outcomes.

There are two ways to present a probability distribution. It can be shown as a probability density function or as a cumulative probability distribution, as shown in Figure 6A and Figure 6B, respectively.

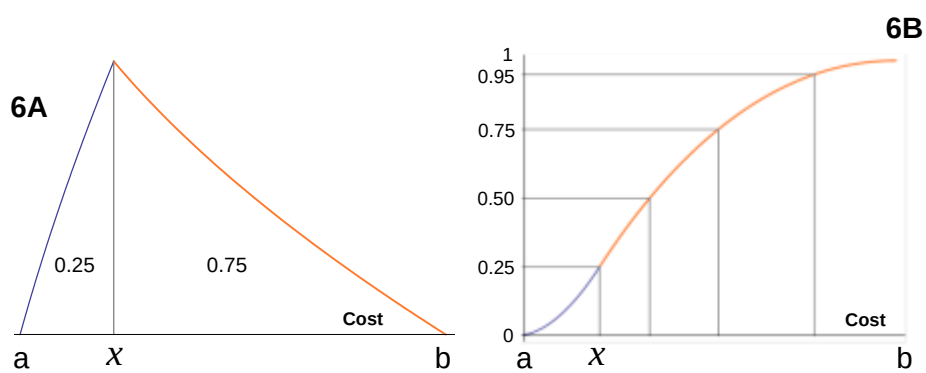


Figure 6. Ways to View a Program's Cost Probability Distribution

In Figure 6, the range of possible cost outcomes for a program is given by the interval $a \le x \le b$. These distributions reveal the confidence that the actual cost of a program will not exceed any cost in the range of possible outcomes. For example, the probability that the actual cost of the program will be less than or equal to *x* is 25 percent. In Figure 6A, this probability is given by the area under the curve. In Figure 6B, this probability is given by the value 0.25 along the vertical axis.

There is an important distinction between the terms ***risk*** and ***uncertainty*** and their use in cost analysis. In general, ***risk*** is the chance of loss or injury. In a situation that includes favorable and unfavorable events, risk is the probability an unfavorable event occurs. In systems engineering risk management, such events might be {failing to achieve performance objectives}, {overrunning the budgeted cost}, or {delivering the system too late to meet user needs}. ***Uncertainty*** is the indefiniteness about the outcome of a situation – it includes favorable and unfavorable events. We analyze uncertainty to measure risk. ***Cost risk*** is a measure of the chance that, due to unfavorable events, the planned or budgeted cost of a program will be exceeded. ***Cost uncertainty*** analysis is the process of measuring the cost impacts of uncertainties associated with a system's technical baseline and cost estimation methodologies.

2.3 MEASURING COST ESTIMATE CONFIDENCE

A cost estimate is stochastic and, therefore, is merely one outcome in a probability distribution of cost outcomes. A cost estimate developed without adjustments for uncertainty is called a point estimate (PE). Thus, a point estimate is just one outcome in a probability distribution of cost outcomes. Measuring the confidence in a point estimate is equivalent to addressing the question "*What is the chance of exceeding the point estimate (or any particular cost) in the range of possible cost outcomes?*"

Identifying and measuring confidence in a point estimate is a fundamental objective of cost risk analysis, especially in the early life cycle phases. Evidence from the community reveals that point estimates are highly uncertain. Recent studies continue to show that, in the early development milestones, the confidence that the final cost of a program falls below its point estimate is less than 50 percent [Garvey, 2012]. Basing or planning a program's budget on the point estimate alone is a high-risk decision. GAO found that "*budgeting programs to a risk-adjusted point estimate that reflects a program's risks is critical to its success in achieving objectives*" [GAO, 2009]. GAO further observed that overly optimistic assumptions and unrealistic expectations are significant factors for cost growth above point estimates.

Developing a point estimate is traditionally done from a work breakdown structure (WBS). Shown in Figure 7 and Figure 8, a WBS is a hierarchical framework that depicts all elements of cost associated with the tasks and activities needed to acquire a program or system.

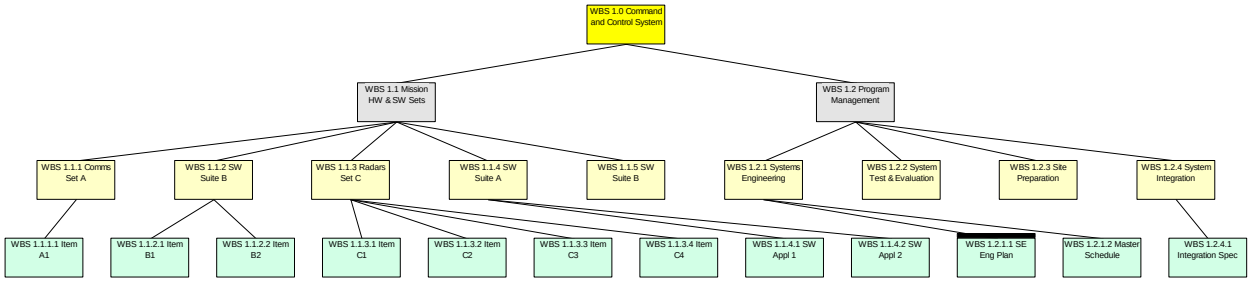


Figure 7. A WBS is a Hierarchy of Cost Elements

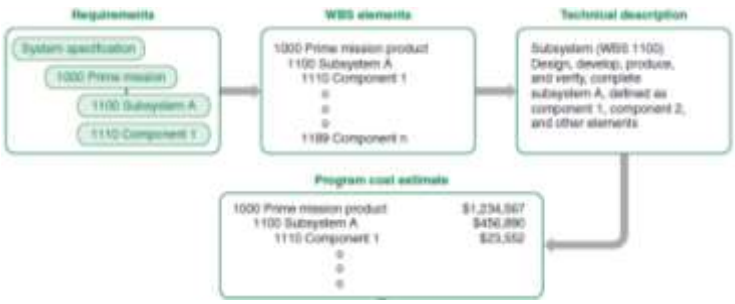


Figure 8. A WBS Integrates Requirements, Technical, and Cost Perspectives[15]

Work breakdown structures can be complex. They may involve many segments and levels, as well as numerous cost elements. Work breakdown structures are unique to the system under consideration and the program's life cycle phase. They are developed according to the specific requirements and functions the system has to perform.

The WBS is the definitive cost element structure and cost model of the program, where the summation of its element costs across WBS levels forms an estimate of total program cost. Initially, this estimate is usually the point estimate cost. In a similar way, the WBS serves as a cost risk model of the program, where the summation of its element cost ranges across WBS levels forms a probability distribution of possible total cost outcomes, one of which is the point estimate. Consider Figure 9.

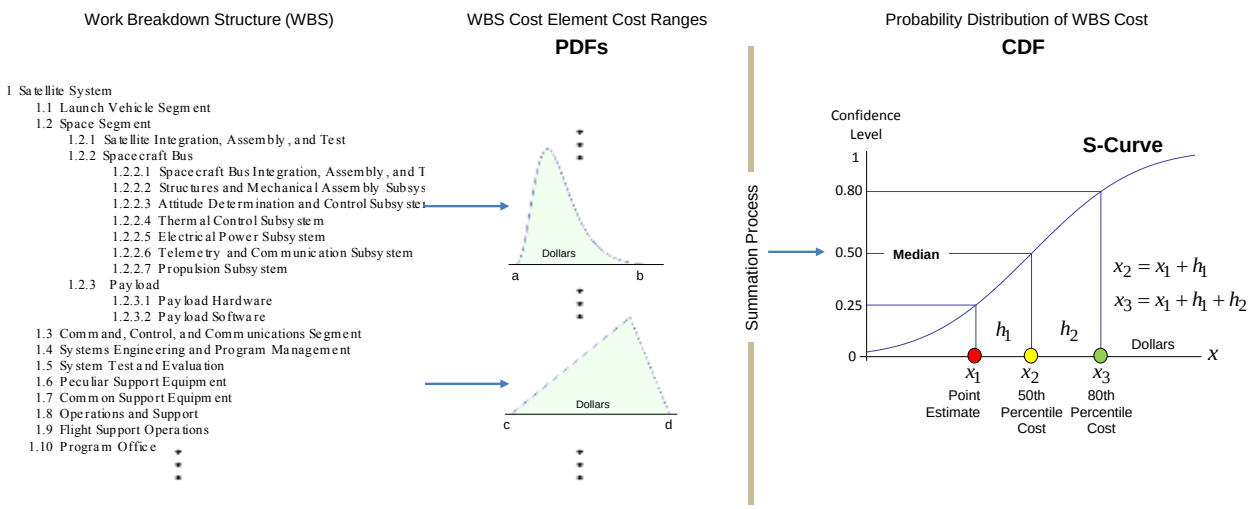


Figure 9. Measuring Cost Estimate Confidence with Probability Distributions

Figure 9 illustrates using the WBS as a cost risk model. Within this, analysts develop ranges around individual cost equation parameters or cost elements. These ranges are stochastically summed to produce a probability distribution of $Cost_{WBS}$, shown on the right of Figure 9. The vertical axis of the distribution provides the measures of cost estimate confidence. The confidence level of the point estimate can be found by locating where it falls in the range of other possible cost outcomes. In Figure 9, the point estimate has a 25 percent confidence level. This means there is a 75 percent chance the actual cost will exceed the point estimate due to risks identified and quantified in the manner described. In Figure 9, values to the right of x_1 are other possible cost outcomes along with their associated confidence levels.

[15] Figure Source: GAO Cost Estimating and Assessment Guide: Best Practices for Developing and Managing Capital Program Costs, GAO-09-3SP, March 2009.

The PDF is the most common form of a probability distribution used to characterize the cost uncertainties of elements that comprise a program's work breakdown structure (WBS). This is illustrated in Figure 9 by the elements shown on the left, which is the input side of a cost risk analysis. The right side of Figure 9 shows the outputs of a cost risk analysis, where the CDF or S-curve is the most common form used to express percentile levels of confidence that the actual cost of a program is less than or equal to a value *x*.

The S-curve in Figure 9 provides decision-makers an analytical basis for tradeoffs between a program's point estimate cost and its confidence. For example, if a program's point estimate cost is x_1 dollars, then the S-curve reveals the amount of additional dollars needed to plan or budget the program at a desired or specified level of confidence. Cost analysts can use the range of possible costs revealed by the S-curve to signal a review of the major cost risk drivers responsible for its variance, before settling too soon on a cost confidence level to plan or budget the program. Furthermore, the results presented by an S-curve can spark a series of design options taken to reduce risk evident in the variance.

Practice Point 5: In Figure 9, x_1 denotes a program's point estimate cost. The difference between x_1 and cost outcomes greater than x_1 is the amount of risk dollars implicit in that cost. For instance, if a program is budgeted to the 50th percentile cost x_2, then relative to x_1 there are h_1 risk dollars contained in x_2. If a program is budgeted to the 80th percentile cost x_3, then relative to x_1 there are $(h_1 + h_2)$ risk dollars contained in x_3. Although there are more risk dollars contained in x_3 than in x_2, there is less chance of a program cost overrun at the 80th percentile confidence level than at the 50th percentile – where an overrun has even odds of occurring. Analyzing the probability distribution of $Cost_{WBS}$ in this way provides an otherwise unseen tradeoff between cost risk dollars and cost estimate confidence.

2.4 COST RISK ANALYSIS PROCESS: FUNDAMENTAL STEPS

Mentioned earlier, cost risk analysis is an inseparable part of cost analysis. Many different elements of a system's technical baseline and cost estimate are involved, as well as individual experts. Recently, the cost analysis community developed steps to guide the conduct of a cost risk analysis. They are summarized below and were excerpted from the referenced report [GAO, 2009].

- Determine the program cost drivers and associated risks;
- Develop probability distributions to model various types of uncertainty;
- Account for correlation between WBS element costs to properly capture cost risk;
- Perform the uncertainty analysis using a Monte Carlo simulation or other analytical approaches;
- Identify the probability level associated with the point estimate;
- Recommend a program budget sufficient to achieve targeted levels of confidence;
- Allocate, phase, and convert a risk-adjusted cost estimate to then-year dollars and identify high-risk elements to help in risk mitigation efforts.

Along with these steps, the community published guidelines in [GAO, 2009] to aid in the identification of potential sources of program cost estimate uncertainty. They are presented in Table 2.

Uncertainty	Definition	Example
Business or economic	Variations from change in business or economic assumptions	Changes in labor rate assumptions—e.g., wages, overhead, general and administrative cost—supplier viability, inflation indexes, multiyear savings assumptions, market conditions, and competitive environment for future procurements
Cost estimating	Variations in the cost estimate despite a fixed configuration baseline	Errors in historical data and cost estimating relationships, variation associated with input parameters, errors with analogies and data limitations, data extrapolation errors, optimistic learning and rate curve assumptions, using the wrong estimating technique, omission or lack of data, misinterpretation of data, incorrect escalation factors, overoptimism in contractor capabilities, optimistic savings associated with new ways of doing business, inadequate time to develop a cost estimate
Program	Risks outside the program office control	Program decisions made at higher levels of authority, indirect events that adversely affect a program, directed funding cuts, multiple contractor teams, conflicting schedules and workload, lack of resources, organizational interface issues, lack of user input when developing requirements, personnel management issues, organization's ability to accept change, other program dependencies
Requirements	Variations in the cost estimate caused by change in the configuration baseline from unforeseen design shifts	Changes in system architecture (especially for system of systems programs), specifications, hardware and software requirements, deployment strategy, critical assumptions, program threat levels, procurement quantities, network security, data confidentiality
Schedule	Any event that changes the schedule: stretching it out may increase funding requirements, delay delivery, and reduce mission benefits	Amount of concurrent development, changes in configuration, delayed milestone approval, testing failures requiring rework, infeasible schedule with no margin, overly optimistic task durations, unnecessary activities, omission of critical reviews
Software	Cost growth from overly optimistic assumptions about software development	Underestimated software sizing, overly optimistic software productivity, optimistic savings associated with using commercial off-the-shelf software, underestimated integration effort, lack of commercial software documentation, underestimating the amount of glue code needed, configuration changes required to support commercial software upgrades, changes in licensing fees, lack of support for older software versions, lack of interface specification, lack of software metrics, low staff capability with development language and platform, underestimating software defects
Technology	Variations from problems associated with technology maturity or availability	Uncertainty associated with unproven technology, obsolete parts, optimistic hardware or software heritage assumptions, feasibility of producing large technology leaps, relying on lower reliability components, design errors or omissions

Table 2. Potential Areas of Program Cost Estimate Uncertainty[16]

[16] Table Source: GAO Cost Estimating and Assessment Guide: Best Practices for Developing and Managing Capital Program Costs, GAO-09-3SP, March 2009.

3.0 TWO WORLDS: MONTE CARLO SIMULATION AND METHOD OF MOMENTS

This section presents the two primary domains where cost risk analysis methods fall. They are Monte Carlo simulation techniques and non-simulation approaches categorized by a class of procedures known as method of moments. This discussion introduces these approaches, highlights differences, and provides guidelines on their use. Popular industry tools that execute these approaches are identified and nuances on their application are explained.

This section also identifies and clarifies issues with certain technical topics that arise in cost risk analysis. These include capturing correlation between WBS element costs, use of normal or lognormal distributions to measure confidence intervals around point estimates, and sample sizes for Monte Carlo simulations.

3.1 THE MONTE CARLO METHOD

The Monte Carlo method is a random sampling technique that empirically derives numerically feasible solutions to a mathematical problem. The technique is best applied to problems not amenable to closed form solutions derived by deterministic algebraic methods.

The Monte Carlo method falls into a class of techniques known as simulation. Simulation has varying definitions among practitioners. For instance, Winston (1994) defines simulation as a technique that imitates the operation of a real world system as it evolves over time. Rubinstein (1981) offers the following: "simulation is a numerical technique for conducting experiments on a computer, which involves certain types of mathematical and logical models that describe the behavior of a business or economic system over extended periods of real time."

The Monte Carlo method involves the generation of random samples from known or assumed probability distributions. The process of generating random samples from distributions is known as random variate generation or Monte Carlo sampling. Simulations driven by Monte Carlo sampling are Monte Carlo simulations. One of the earliest applications of Monte Carlo simulation to cost analysis was at the RAND Corporation [Dienemann, 1966]. Since then, Monte Carlo simulation has become (and remains) a popular approach for modeling and measuring cost uncertainty.

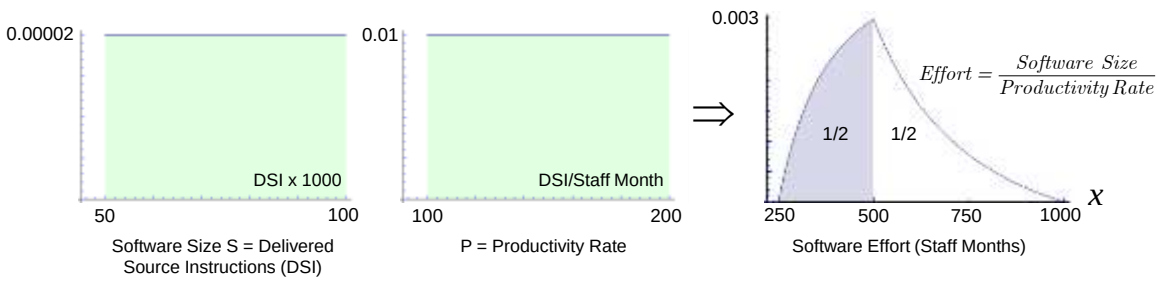
For cost risk analysis, Monte Carlo simulation is typically applied to a work breakdown structure to develop an empirical distribution of program cost. The WBS serves as the cost model of the program within which to conduct the simulation. The steps in a Monte Carlo simulation are as follows:

- For each variable in the WBS whose value is uncertain, randomly select one value from the probability distribution that characterizes its uncertainty (e.g. from a uniform or triangle distribution).
- Once a set of possible values for each uncertain variable has been randomly drawn, combine them according to the cost estimation relationships specified in the WBS. This process produces a single randomly generated value for $Cost_{WBS}$, which denotes the work breakdown structure's total cost.
- The above steps are repeated thousands of times producing thousands of values of $Cost_{WBS}$. Each value represents one of these thousands of possible outcomes for $Cost_{WBS}$.
- From the above step, develop a frequency distribution of the outcomes for $Cost_{WBS}$. This is the empirical probability distribution of $Cost_{WBS}$. It is an approximation to the true (but unknown) underlying distribution of $Cost_{WBS}$, formed by the Monte Carlo simulation running through the WBS.

In cost risk analysis, Monte Carlo simulations are generally static simulations. Static simulations are used to study behavior at a discrete point in time. To illustrate a static simulation, and Monte Carlo sampling, consider the problem of determining the average effort (staff months) to develop a software application. Assume *Effort* is computed by the cost estimation relationship

$$Effort = \frac{Software\ Size}{Productivity\ Rate} = \frac{S}{P} \tag{1}$$

where uncertainties in S and P are given by the uniform distributions shown on the left side of Figure 10. In the Monte Carlo method, values of S and P are randomly sampled from their distribution functions. With them, a value for *Effort* is then computed according to Equation 1. This process is repeated thousands of times to produce a static simulated distribution of *Effort*. The simulated distribution is an empirical approximation of the exact probability distribution[17] of *Effort*, shown on the right of Figure 10.



Software Size S = Delivered Source Instructions (DSI)	P = Productivity Rate	Software Effort (Staff Months)
53,423	113	472.77
97,346	182	534.87
78,321	142	551.56
92,837	187	496.45
51,837	158	328.08
93,283	163	572.29
73,176	109	671.34
83,710	151	554.37
67,194	169	397.60
66,512	102	652.08

10 Random Samples: Estimated Mean 523.14 Staff Months
Exact Mean (Derived Expected Value) 519.86 Staff Months

Figure 10. Monte Carlo Sampling: 10 Random Samples from the Distributions of S and P

In Figure 10, ten random samples of S and P are shown. For each S and P pair, the corresponding value for *Effort* is computed according to Equation 1. This produces ten random samples (outcomes) of *Effort*, listed on the right side of Figure 10. The average of all ten sampled values of *Effort* is 523.14 staff months. This is an empirical approximation of the exact mean of *Effort*, shown in Figure 10 as 519.86 staff months. The percentage error between these two values is only 0.63 percent. Increasing the number of random samples of S and P would further reduce this percentage error. If 100,000 random samples of S and P were drawn, then the percentage error between the simulated mean and the exact mean of *Effort* is reduced to 0.11 percent. Figure 11 shows the negligible difference between the empirical probability distribution (shown by the points) and the exact probability distribution (shown by the solid line) of *Effort* using 100,000 random samples of S and P. For those interested, the exact probability distribution of *Effort* in Figure 10 is given by Equation 1A and shown by the solid line in Figures 11 and 12.

$$Prob(Effort \le x) = \begin{cases} \left(\frac{250}{x} + \frac{x}{250}\right) - 2 & \text{if } 250 \le x \le 500 \\ 3 - \left(\frac{1000}{x} + \frac{x}{1000}\right) & \text{if } 500 \le x \le 1000 \end{cases} \tag{1A}$$

[17] The probability distribution of a combination of random variables can be difficult and, in some cases, intractable to derive in an exact algebraic form (as in Equation 1A). Monte Carlo simulation will always generate an empirical approximation to the exact form of the distribution; hence, in practice, simulation is often the most efficacious approach. Figure 11 illustrates this discussion.

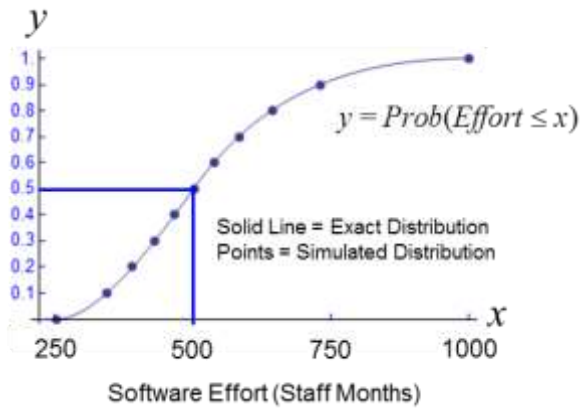


Figure 11. Comparing a Monte Carlo Simulation to the Exact Distribution of *Effort*

Figure 10 and Figure 11 illustrate ways to present a probability distribution, discussed in Section 2.2. Figure 12A is the probability density function (PDF) of *Effort*. Figure 12B is the cumulative distribution function (CDF) for *Effort*, informally called the S-curve. In Figure 12, the range of possible values of *Effort* is the interval $250 \le x \le 1000$. The PDF or CDF reveal the confidence of not exceeding any value in the range of possible values. For example, the probability that the true software development effort will be less than or equal to x = 500 staff months is 50 percent. In Figure 12A, this probability is given by the area under the curve. In Figure 12B, this probability is given by y = 0.5 along the vertical axis.

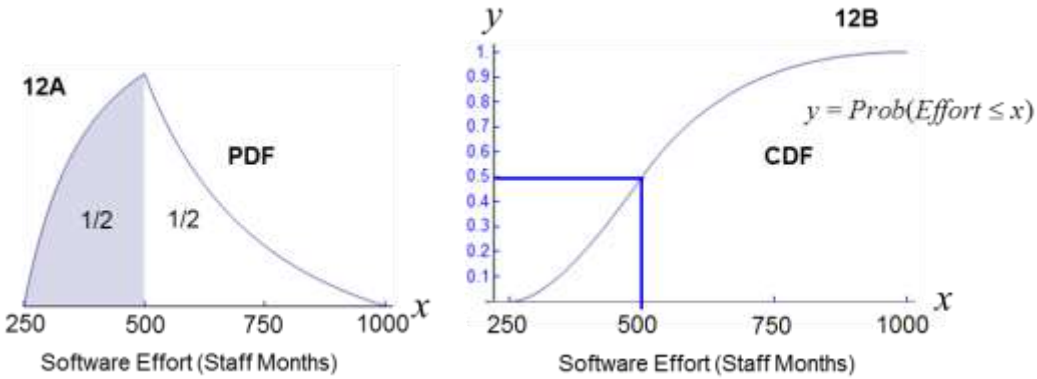


Figure 12. The PDF and CDF of *Effort* in Figure 10

Figure 12 illustrates a property about the mean, median, and mode of a probability distribution. They are equal when the distribution is normal; however, because the distribution in Figure 12 is skewed the mean, median, and mode are not equal. Here, the median and the mode are each 500 staff months but the exact mean of this distribution is 519.86 staff months. In this case, the mean is larger than the median or the mode because the probability distribution of *Effort* is skewed to the right[18]. Refer to Garvey (2000) for a complete analytical derivation of the exact forms of the PDF and CDF in Figure 12.

[18] The mode of a distribution is the value of x at the peak of the distribution. In Figure 12, this occurs at x = 500 staff months. The median of a distribution is the value of x that occurs at exactly the 50th percentile. In Figure 12, this happens to occur at x = 500 as shown by 1/2 the area under the PDF or by the value y = 0.50 along the vertical axis of the CDF. In general, the mean, median, and mode of a probability distribution are not always the same.

3.1.1 Sample Size for Monte Carlo Simulations

In Monte Carlo simulations, a question often asked is "*How many trials (random samples) are needed to have confidence in the outputs of the Monte Carlo simulation?*" The statistical literature provides guidelines for determining sample size as a function of the precision desired in the outputs of a simulation. Specifically, the formulas below address the question: "*What sample size is needed so that with probability α the true value of an uncertain variable falls between a pair of values generated by the Monte Carlo simulation?*" In Figure 6A, α is shown by the area under the curve. In Figure 6B, α is shown by the values along the vertical axis. Figure 6B is most commonly used in cost risk analysis to show cost estimate confidence and the range of possible program cost outcomes.

***Morgan-Henrion Guideline* (1990)**: Define m as the sample size for the Monte Carlo simulation. Let x_α be the α-fractile of the random variable X; that is, $Prob(X \le x_\alpha) = \alpha$. Let c satisfy the probability $P(-c \le Z \le c) = \alpha$, where $Z \sim N(0,1)$ is the standard normal probability distribution. Then, the pair of fractiles $(\hat{x}_i, \hat{x}_k)$ generated by the Monte Carlo simulation, with

$$i = \alpha - c\sqrt{\frac{\alpha(1-\alpha)}{m}} \tag{2}$$

$$k = \alpha + c\sqrt{\frac{\alpha(1-\alpha)}{m}} \tag{3}$$

contains x_α with probability α. For different sample sizes, Figure 13 shows with probability $\alpha = 0.95$ (or $c = 2$) the values of i and k such that the true median of the distribution falls between $(\hat{x}_i, \hat{x}_k)$. The lower and upper curves in Figure 13 are generated from Equations 2 and 3, respectively. As the sample size m increases, the difference between these curves decreases dramatically. With 100 samples, the true median value $x_{0.50}$ of the random variable X falls between the pair of fractiles $\hat{x}_{0.40}$ and $\hat{x}_{0.60}$, generated by the Monte Carlo simulation, with probability 0.95. Increasing that sample size by a factor of 100 brings the same degree of confidence to be within $\hat{x}_{0.49}$ to $\hat{x}_{0.51}$.

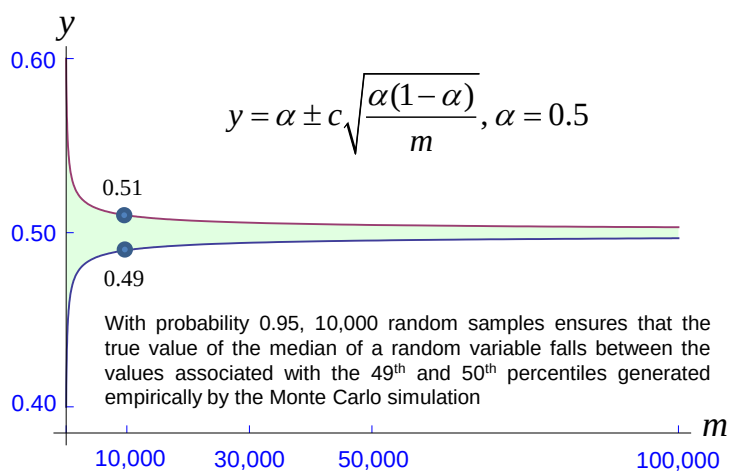


Figure 13. Number of Trials for Monte Carlo Simulations

Practice Point 6: 10,000 trials (or random samples) are generally sufficient to meet the precision requirements for Monte Carlo simulations for cost risk analysis.

3.1.2 Monte Carlo Simulation Applied to Work Breakdown Structures

Monte Carlo simulation is commonly applied to a work breakdown structure (WBS) when it is used to derive a probability distribution of a program's cost risk. Mentioned in Section 2.3, the WBS is the definitive cost element structure and cost model of a program, where the summation of element costs across WBS levels forms an estimate of total cost. In a similar way, the WBS serves as a cost risk model of the program, where the summation of element cost ranges across WBS levels forms a probability distribution of possible total cost outcomes, one of which is the point estimate. This was illustrated in Figure 9. For convenience, it is shown below as Figure 14. The following illustrates a Monte Carlo simulation applied to a WBS consisting of five cost elements.

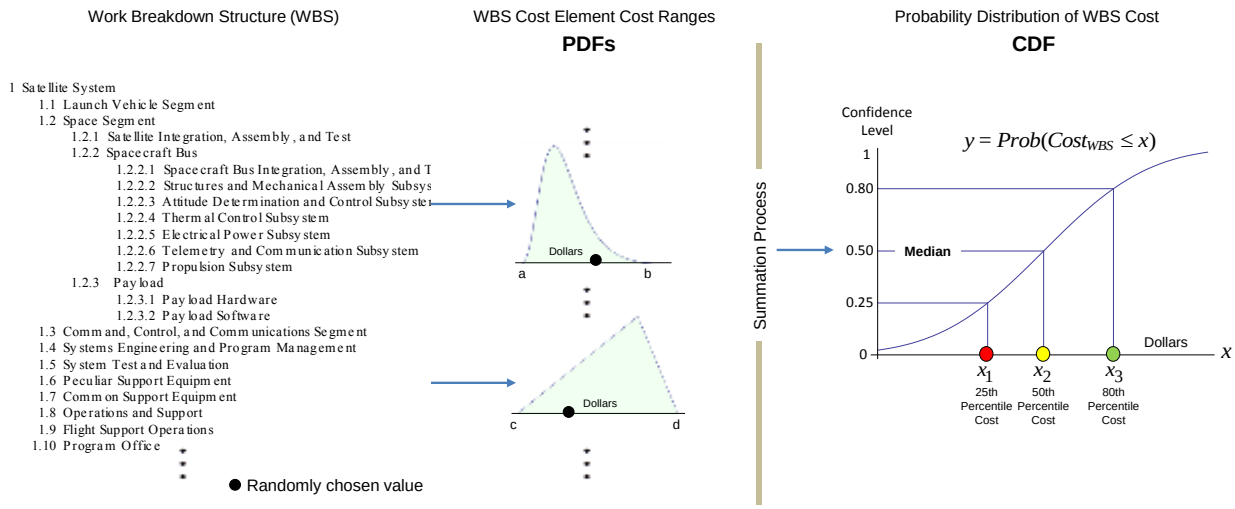


Figure 14. Monte Carlo Simulation Applied to a Work Breakdown Structure

Figure 15 presents a simple work breakdown structure consisting of five cost elements X1, X2, X3, X4, and X5. A point estimate cost for each element is shown, along with an uncertainty distribution around each estimate. For each WBS element, a random value from its cost uncertainty distribution is taken. These randomly selected values are then summed to form one estimate of total cost. This process is repeated thousands of times (e.g., 10,000 times or more) to produce an empirically derived overall probability distribution of total cost. This is the Monte Carlo simulation process, with each colored circle in Figure 15 representing a single thread or single pass through one of the thousands of Monte Carlo samples.

Mentioned earlier, the outcome of this process is a frequency distribution derived from these *n* sampled values. This distribution is the Monte Carlo simulated distribution of total cost. Figure 16 presents the results of 10,000 Monte Carlo samples of the WBS in Figure 15. It shows the resultant probability distribution of the WBS total cost, given by Equation 4.

$$Cost_{WBS} = X1 + X2 + X3 + X4 + X5 \qquad (4)$$

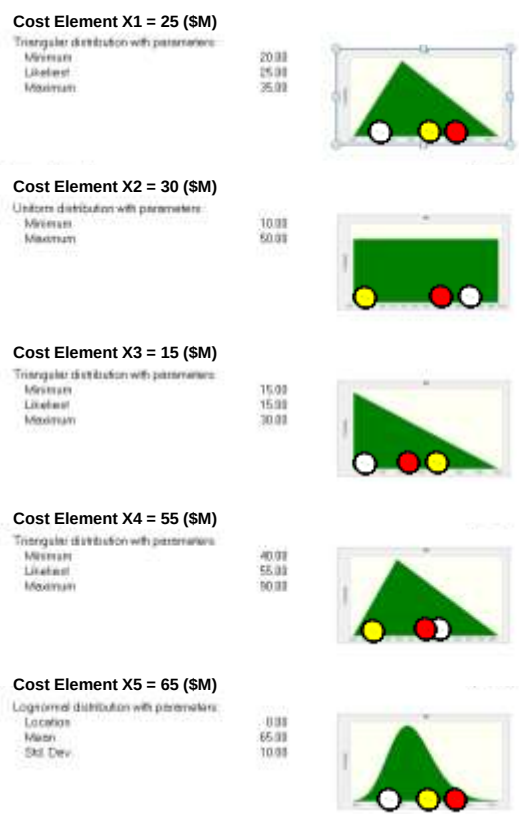


Figure 15. A WBS for Monte Carlo Simulation

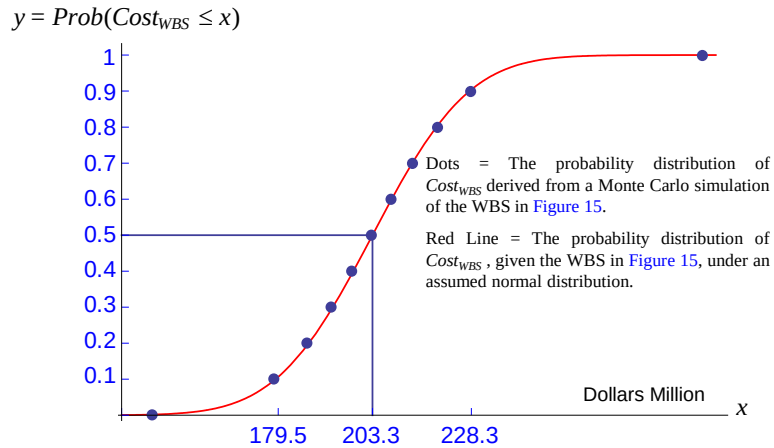


Figure 16. Probability Distribution of $Cost_{WBS}$

In Figure 16, the dots are the probability distribution $Cost_{WBS}$ generated by the Monte Carlo simulation of the WBS in Figure 15. The simulation produced a total cost mean μ of \$203.3M and a standard deviation σ of \$19M. In Figure 16, the solid red line is the probability distribution of $Cost_{WBS}$ assuming its possible cost outcomes fall along a normal probability distribution – with mean and variance given by the sums of the means and variances of the WBS element cost ranges in Figure 15.

Observe the closeness of the simulated and normal probability distributions in Figure 16. There are reasons for this result. One reason is due to the assumed mutual independence between WBS element costs X1, X2, X3, X4, and X5 given in Figure 15. With this, the famous Central Limit Theorem (CLT) enters the picture and ensures the eventual tendency of the simulated distribution to approach a normal distribution[19]. This tendency is affected by many factors, which will be discussed in Section 3.2.

3.1.3 Correlation and Monte Carlo Simulations

For years, the cost analysis community has addressed the topic of correlation and how to represent it in a cost risk analysis. Despite a large body of technical work on this topic, the community needs practical guidance consistent with the subtleties of statistical theory. Given this, this section offers a practical approach for modeling correlation and capturing its effects on program cost. The approach is presented in the context of Monte Carlo simulations, but it can easily be woven into other cost risk analysis methods. Illustrative discussions are provided herein and in Appendix A; first, some background.

What Correlation is and Why it Matters

Correlation ρ is a statistical measure of the "co-variation" between two random variables. It measures the strength and direction of change in one random variable with change in another random variable. Regarding strength, correlation is a continuous measure whose magnitude ranges between negative one and positive one. Regarding direction, correlation can be positive or negative. *Positive correlations* fall in the interval $0 < \rho \leq 1$. *Negative correlations* fall in the interval $-1 \leq \rho < 0$. Uncorrelated random variables have correlation $\rho = 0$.

[19] Informally, the Central Limit Theorem (CLT) establishes the fundamental result that the sum of independent identically distributed random variables with finite variance approaches a normally distributed random variable as their number increases; in particular, if enough random samples (e.g., Monte Carlo samples) are repeatedly drawn from any distribution, the sum of the sample values can be thought of, approximately, as an outcome from a normally distributed random variable. The requirement for identically distributed random variables has been relaxed in modern variants of the original CLT. Refer to Garvey (2000) for an extensive discussion of the CLT and conditions for its applicability in cost risk analysis.

From a cost analysis perspective, the strength of correlations between a program's WBS element costs affects the magnitude of the its overall cost risk, as measured by the variance or standard deviation of the total cost probability distribution[20]. The positive or negative direction of correlation affects whether this magnitude increases or decreases cost risk, respectively.

Thus, correlation is a required consideration in modeling the cost uncertainty of a program. Ignoring correlation is equivalent to setting its value to zero. Doing this when there is actual positive correlation between one or more pairs of WBS element costs can significantly underestimate a program's "true" cost risk. The extent of the potential underestimation is shown in Figure 17.

Figure 17 shows the percent that cost risk σ is underestimated if correlation was assumed equal to zero between all pairs of WBS element costs, instead of capturing an actual positive constant correlation between all pairs. For example, suppose $\rho = 0$ was initially assumed between all WBS element costs in a 10 element WBS. If it later became evident that $\rho = 0.40$ between all WBS element costs, then the initial measure of cost risk is underestimated by 53 percent. In a 30 element WBS, the initial measure of cost risk would be underestimated by 72 percent. Seen in Figure 17, the underestimation of cost risk worsens exponentially as the number of WBS cost elements increases.

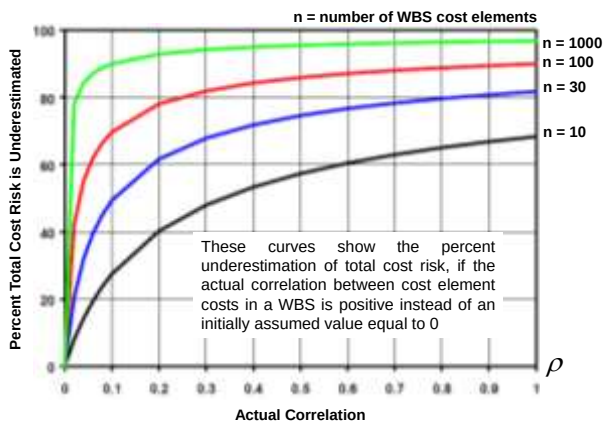


Figure 17. Possible Impacts of Positive Correlation on Cost Risk [Book, 1999]

Why and How Correlation Enters the Scene

The cost of a program derives from a summation process when it is based on a work breakdown structure; that is, a program's total cost is the sum of its work breakdown structure element costs. Furthermore, the statistical mean of the program's total cost is the sum of the statistical means of its WBS element costs. However, the statistical variance of the program's total cost is not the sum of the statistical variances of its WBS element costs. To see why, we need only look at the sum of two random variables *X* and *Y*. What is the variance of their sum? The answer is given by Equation 5.

$$Var(X+Y) = \sigma^2_{X+Y} = Var(X) + Var(Y) + 2\rho_{X,Y}\sigma_X\sigma_Y \tag{5}$$

The variance of $(X+Y)$ is not just the sum of the variance of *X* plus the variance of *Y*. The last term in Equation 5 is called the covariance (or the "co-variation") between *X* and *Y*. In Equation 5, $\rho_{X,Y}$ is the Pearson product-moment correlation between *X* and *Y* and $\sigma_X\sigma_Y$ is the product of their respective standard deviations. This is technically why and how correlation enters the scene. The Monte Carlo simulation of the work breakdown structure in Figure 15 reflected the condition that all pairs of WBS

[20] The standard deviation σ is the square root of the variance σ^2 of a random variable.

element costs were uncorrelated – a change in the cost of one element was not associated with a change in the cost of another element. However, this condition is not the common case. In work breakdown structures more complex than the one given in Figure 15, correlation is often found between many pairs of WBS element costs. If the presence of correlation is such that an increase in the cost of one element is associated with an increase in the cost of another element, then this positive correlation causes the variance of the sum of these WBS element costs to increase. The reason for this is seen in Equation 6.

$$Var(X+Y) = \sigma^2_{X+Y} = \begin{cases} Var(X) + Var(Y) & \text{if } \rho_{X,Y} = 0 \\ Var(X) + Var(Y) + \underbrace{2\rho_{X,Y}\sigma_X\sigma_Y}_{Covariance(X,Y)} & \text{otherwise} \end{cases} \tag{6}$$

The variance of X + Y increases by this amount, which is the covariance between X and Y, when Pearson product-moment correlation is present between them

Equation 6 is the variance of the sum of just two random variables X and Y if $\rho_{X,Y} = 0$ (X and Y are uncorrelated) or if $\rho_{X,Y} \neq 0$ (X and Y are correlated). The term $2\rho_{X,Y}\sigma_X\sigma_Y$ is the covariance of X and Y. In Equation 6, the covariance becomes larger and larger as more and more random variables with positive correlations are summed. An example is shown by the equation below.

$$Var(X+Y+Z) = \sigma^2_{X+Y+Z} = \begin{cases} Var(X) + Var(Y) + Var(Z) & \text{if } \rho_{X,Y}, \rho_{X,Z}, \rho_{Y,Z} = 0 \\ Var(X) + Var(Y) + Var(Z) + \underbrace{2\rho_{X,Y}\sigma_X\sigma_Y + 2\rho_{X,Z}\sigma_X\sigma_Z + 2\rho_{Y,Z}\sigma_Y\sigma_Z}_{Covariance(X,Y,Z)} & \text{otherwise} \end{cases}$$

The variance of X + Y + Z increases by this amount, which is the covariance between X, Y, and Z, when Pearson product-moment correlation is mutually present between them

Thus, correlation between WBS element costs in a WBS can have significant effects on the magnitude of cost risk, given by the variance σ^2 or the standard deviation σ of the total cost probability distribution.

Pearson and Spearman Rank Correlation Measures

In Equation 6, the correlation coefficient $\rho_{X,Y}$ is the Pearson product-moment correlation. Pearson's correlation measures the strength and direction of the linearity between X and Y. It is the <u>*only correct*</u> correlation measure for summing the variances of random variables and, in our context, when summing the variances of a program's WBS element costs.

There are other types of correlation measures in statistics. One is Spearman's rank correlation. Rank correlation is also measured in the interval $-1 \leq \rho_{rank} \leq 1$. It measures the strength and direction of the monotonicity between two random variables. Monotonicity and linearity can be different behaviors between pairs of random variables. Thus, Spearman's rank correlation and Pearson's product-moment correlation are not guaranteed to produce the same measures. The following illustrates this point.

In Figure 18, suppose Y is a random variable that is a function of X. Suppose X is a random variable whose outcomes are uniformly distributed in the interval 0 to 1. In Figure 18, the function on the left has a Pearson correlation of $\rho_{X,Y} = 0.8732$ and a rank correlation equal to one ($\rho_{rank} = 1$). The function on the right has a Pearson correlation of $\rho_{X,Y} = 0.7861$ and a rank correlation equal to one ($\rho_{rank} = 1$). Why are the Pearson and rank correlations so different? The answer is because Y is a perfectly monotonically increasing function of X over the indicated domains. Rank correlation measures the strength and direction of monotonicity *only*. A function Y can have a perfect rank correlation *but a very different Pearson correlation*. This is because a monotonic function need not be a linear function.

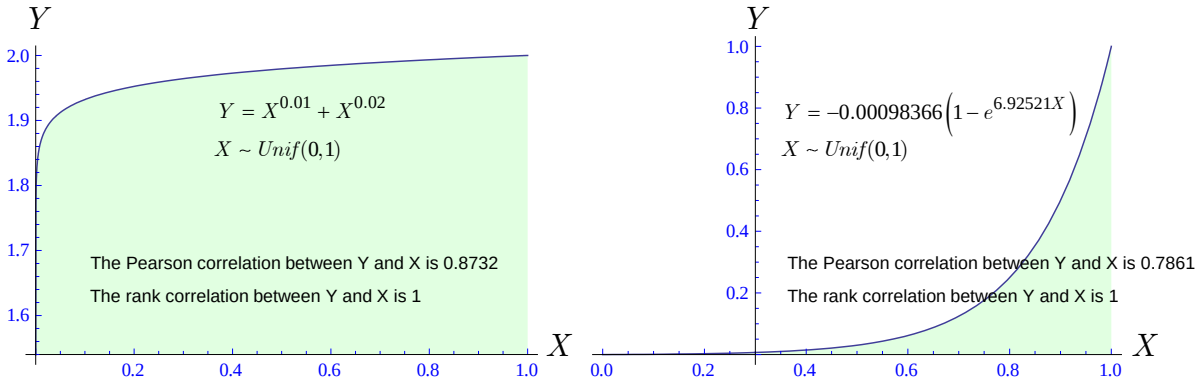


Figure 18. Comparing Pearson and Rank Correlation Measures

Guidance on Capturing Correlation in Monte Carlo Simulations

This discussion provides implementation guidance on capturing Pearson correlation when a WBS Monte Carlo simulation is used to a conduct cost risk analysis. A post-simulation adjustment technique (PSAT) is described. It is premised on the following: (1) correlation cannot be ignored – doing so is equivalent to setting it equal to zero (2) Pearson correlation *must only be used* and modeled correctly in a WBS to avoid double counting correlation, or mixing different types of correlation. The latter is aimed at preventing the generation of invalid measures of cost risk or measures that are unrealistically under- or overestimated.

PSAT operates on three practice realities. The first is that cost analysts often use Monte Carlo simulation spreadsheet technologies to structure and execute cost risk analyses. The second is that within these spreadsheets, there are WBS element cost equations that functionally share some of the same input variables and some that do not. The third is that correlations often exist between many WBS element costs, but too often their values are not empirically known.

In Monte Carlo simulations, including those run with commercial tools such as @Risk or Crystal Ball, Pearson product-moment correlations are automatically captured between functionally related WBS element costs[21]. Informally, this is called *functional correlation*. However, the cost estimation equations throughout a WBS spreadsheet model usually contain a mix of functionally related and non-functionally related element costs. How can Pearson product-moment correlations between certain WBS element costs be captured if they were not functionally defined in the spreadsheet model? The post-simulation adjustment technique was developed to address this issue.

PSAT: A 3-Step Process

A simple model is used to illustrate PSAT. Extending it to cases more complex than the example below is straightforward. Define $Cost_{WBS}$ as a work breakdown structure's total cost. For simplicity, let $Cost_{WBS}$ be the sum of three WBS element cost random variables X_1, X_2, and X_3 given by Equation 7.

$$Cost_{WBS} = X_1 + X_2 + X_3 \tag{7}$$

[21] When conducting a cost risk analysis, do not introduce a rank correlation matrix into a Monte Carlo simulation of a WBS cost model. Rank correlation is not the correct correlation in the variance of the sum of WBS element cost random variables. Pearson product-moment correlations may already be captured in the WBS cost simulation due to functional relationships defined between WBS element costs. Mixing Pearson product-moment correlations with Spearman rank correlations introduced by a rank correlation matrix leads to (a) double counting the effects of correlation on total cost variance and, more importantly, (b) a cost simulation that produces statistical results whose interpretation is unknown.

Suppose the cost of X_2 is a function of the cost X_1 and X_3 but the costs of X_1 and X_3 are estimated separately from each other. From this, it follows that functional correlations exist between the costs of WBS element pairs (X_1, X_2) and (X_2, X_3) but not between the WBS element pair (X_1, X_3). However, despite the absence of functional correlation in the WBS between X_1 and X_3, suppose engineers and analysts know from experience that their costs have some level of positive correlation – which must be captured in the cost risk analysis. How can these correlation considerations be modeled in a Monte Carlo simulation with Pearson product-moment correlations? Mentioned earlier, PSAT was developed for this purpose. The following describes the steps to implement PSAT. Appendix A illustrates PSAT through each implementation step with a numerical example of $Cost_{WBS} = X_1 + X_2 + X_3$.

Step 1. Build and Implement a WBS Monte Carlo Simulation
Build and implement a Monte Carlo simulation within the WBS cost model using the same cost equations and estimating relationships that produced the overall point estimate of total program cost. The simulation will automatically capture all Pearson correlations between WBS element costs that stem from functional relationships between them, as they were defined to produce the original point estimate.

For example, consider the preceding discussion where $Cost_{WBS} = X_1 + X_2 + X_3$. A Monte Carlo simulation of $Cost_{WBS}$ will automatically capture Pearson product-moment correlations between the pair (X_1, X_2) and the pair (X_2, X_3), since they were functionally defined by their cost estimating relationships. Thus, the simulation generates the following statistical mean and variance measures given by Equations 8 and 9, respectively.

$$E(Cost_{WBS})_{Sim} = E(X_1 + X_2 + X_3) = E(X_1) + E(X_2) + E(X_3) \tag{8}$$

$$\begin{aligned} Var(Cost_{WBS})_{Sim} = Var(X_1 + X_2 + X_3) &= Var(X_1) + Var(X_2) + Var(X_3) \\ &\quad + 2\rho_{X_1,X_2}\sigma_{X_1}\sigma_{X_2} + 2\rho_{X_2,X_3}\sigma_{X_2}\sigma_{X_3} \end{aligned} \tag{9}$$

From the preceding, the pair (X_1, X_3) had no functional correlation between them in the WBS cost model. Thus, Equation 9 does not contain the covariance term $2\rho_{X_1,X_3}\sigma_{X_1}\sigma_{X_3}$. The Monte Carlo simulation is effectively operating with the property that $\rho_{X_1,X_3} = 0$, in this case.

Practice Point 7: Do not introduce a rank correlation matrix into this step or into any of the PSAT steps that follow. Mentioned earlier, analysts must not introduce a rank correlation matrix into a Monte Carlo simulation of a WBS cost model. Rank correlation is not the correct correlation in the variance of the sum of WBS element cost random variables. Pearson product-moment correlations may already be captured in the WBS cost simulation due to functional relationships defined between element costs. Mixing Pearson product-moment correlations with Spearman rank correlations introduced by a rank correlation matrix leads to (a) double counting the effects of correlation on total cost variance and, more importantly, (b) a cost simulation that produces statistical results whose interpretation is unknown.

Step 2. Make a Post-Simulation Adjustment to the Simulated Variance
In the discussion about $Cost_{WBS} = X_1 + X_2 + X_3$, it was given that X_1 and X_3 were uncorrelated from a Monte Carlo simulation perspective – they had no functional correlation between them in the WBS cost models. However, suppose engineers and analysts know from experience the costs of WBS elements X_1 and X_3 have some level of positive correlation. This correlation must also be captured in the cost risk analysis. Since correlation only affects $Var(Cost_{WBS})$, a post-simulation adjustment to just Equation 9 is needed. To do this, PSAT defines an "adjusted" variance $adjVar(Cost_{WBS})$ as follows:

$$adjVar(Cost_{WBS}) = Var(Cost_{WBS})_{Sim} + \underbrace{2\rho_{X_1,X_3}\sigma_{X_1}\sigma_{X_3}}_{\text{Additional covariance needed to capture the Pearson product-moment correlation between } X_1 \text{ and } X_3} \quad \text{(10)}$$

$Covariance(X_1, X_3) =$

where $Var(Cost_{WBS})_{Sim}$ comes from Equation 9 which is generated by the Monte Carlo simulation (Step 1), σ_{X_1} is the square root of the $Var(X_1)$ which is generated by the Monte Carlo simulation (Step 1), and σ_{X_3} is the square root of the $Var(X_3)$ which is generated by the Monte Carlo simulation (Step 1). In Equation 10, the Pearson product-moment correlation ρ_{X_1,X_3} for the WBS element pair (X_1, X_3) is a new input. Its value can come from historical data, subject experts, or assigned from the guidelines below.

Guidelines for Assigning Pearson Correlations

The cost analysis community has guidelines for assigning Pearson product-moment correlations in the absence of these measures being derived from functional relationships in a WBS. In Figure 17, observe that the "knee in the curves" occurs in the interval $0.20 \le \rho \le 0.30$. For $\rho > 0.30$, observe there is little change in the percent that cost risk σ is underestimated by not capturing positive correlation when it is present in a work breakdown structure. Choosing a value for ρ in this interval provides most of the needed additional covariance (e.g., required in Equation 10) at the least loss of accuracy [Book, 1999].

Table 3 provides further context when assigning a value to a Pearson product-moment correlation. From a cost analysis perspective, the square of ρ represents the percentage of variation in the cost of one WBS cost element attributable to the influence of variation in the cost of another WBS cost element. If *X* and *Y* are two WBS element costs with $\rho_{X,Y} = 0.32$, then 10 percent of the variation in the cost of cost element *Y* is attributable to the influence of variation in the cost of WBS element *X*. Table 3 provides interpretations of assigned correlations between pairs of WBS element cost random variables, when other ways to derive this measure are absent.

Correlation	% Influenced
0.00	0%
±0.10	1%
±0.32	10%
±0.50	25%
±0.71	50%

Table 3. Percent of Variability in Cost Influenced by Correlation [Book, 1999]

Step 3. Make a Post-Simulation Adjustment to the Simulated Probability Distribution of $Cost_{WBS}$

With Step 2 completed, the simulated probability distribution of $Cost_{WBS}$ must then be adjusted. This is necessary to account for the adjusted variance determined in the preceding step.

For the WBS scenario with $Cost_{WBS} = X_1 + X_2 + X_3$ this adjustment is given by Equation 10. Equation 10 would be tailored accordingly for other scenarios. To adjust the simulated probability distribution of $Cost_{WBS}$, one can reasonably assume the amended distribution is either normal or lognormal with mean $E(Cost_{WBS})_{Sim}$ and variance $adjVar(Cost_{WBS})$. Applying the normal or lognormal distributions in PSAT and in other cost risk analysis contexts is described in the following section. The reader is also directed to Appendix A for a complete numerical illustration of each PSAT step.

3.1.4 Summary

This section introduced Monte Carlo simulation, one of the primary methods of cost risk analysis. The Monte Carlo method is in a class of techniques known as probabilistic modeling. Monte Carlo simulation is a popular approach for modeling and measuring cost uncertainty.

For cost risk analysis, Monte Carlo simulation can easily be used within a work breakdown structure to develop the empirical distribution of a program's total cost. The WBS serves as the mathematical model of the cost of the program within which to build and conduct the simulation. The following summarizes key practice points about Monte Carlo simulation when it is the protocol for conducting cost risk analysis.

Practice Point 8: Monte Carlo Simulations Capture Pearson Functional Correlations

Monte Carlo simulations will automatically capture Pearson product-moment correlations that may be present between WBS element costs by virtue of the way analysts define their equations (or relationships) in a cost model.

Practice Point 9: Do Not Introduce Rank Correlations into a Cost Risk Analysis

Given Practice Point 8, do not introduce a rank correlation matrix into a Monte Carlo simulation of a WBS cost model. Rank correlation is not the correct correlation in the variance of the sum of WBS element cost random variables. Pearson product-moment correlations may already be captured in the WBS cost simulation due to functional relationships defined between element costs. Mixing Pearson product-moment correlations with Spearman rank correlations introduced by a rank correlation matrix leads to (a) double counting the effects of correlation on total cost variance and, more importantly, (b) a cost simulation that produces statistical results whose interpretation is unknown.

Practice Point 10: PSAT Applies to Correlating Random Variables After Simulations

Given Practice Point 8, apply the Post-Simulation Adjustment Technique (PSAT) to those WBS element costs that indeed co-vary but, for a variety of reasons (e.g., no functional relationships between them) their correlations are not automatically captured in the Monte Carlo simulation.

Practice Point 11: PSAT Adjusts the Simulated Variance of Total Cost

The Post-Simulation Adjustment Technique only adjusts the simulated variance of a sum of WBS element costs. The mean of the sum of WBS element costs is unaffected by correlation. The Post-Simulation Adjustment Technique only requires analysts to specify those correlations not automatically captured by the simulation. All other inputs required for PSAT are produced by the simulation.

Practice Point 12: Guidance on Assigning Correlations

Given Practice Point 11, if analysts need to assign values to correlations not automatically captured in a Monte Carlo simulation, then do so in accordance with the guidelines in Table 3 and Figure 17. Realize that certain types of probability distributions cannot be positively correlated at the maximum value of 1 or negatively correlated at the minimum value of -1 in the correlation interval $-1 \le \rho \le 1$ [Garvey, 2000]. Therefore, caution is needed to avoid assigning an infeasible correlation between the costs of pairs of WBS elements.

Cost risk analysis involves judgment, experience, and knowledge of statistical science. It is better to incorporate some level of correlation between WBS element costs known to co-vary than ignoring their associations altogether. Sensitivity analyses should always be conducted around assigned correlation values to examine the reasonableness of their effects on a program's overall measure of cost risk.

3.2 THE METHOD OF MOMENTS

The method of moments is a procedure in classical statistics to estimate the unknown mean and variance of a population by random samples from the population. Inferences from the sample mean and sample variance are used to make inferences about the population mean and population variance. A *moment* is a statistical term referring to central tendency measures of a random variable or its probability distribution. The mean is the first moment. The variance is a function of the second moment of a random variable.

In cost risk analysis, the method of moments (MOM) refers to deriving the mean and variance of the cost of a program as functions of the means and variances of the costs of its WBS cost elements. From these measures, the probability distribution of $Cost_{WBS}$ is formed. The method of moments produces analytically derived measures of cost risk, while Monte Carlo simulation empirically derives them from thousands of random trials or samples.

3.2.1 Method of Moments: Applied to Work Breakdown Structures

The method of moments is commonly applied to a work breakdown structure (WBS) when it is used to derive the probability distribution of $Cost_{WBS}$. Mentioned in Section 2.3, the WBS is the definitive cost element structure and cost model of a program, where the summation of WBS element costs across WBS levels forms an estimate of total program cost. Similarly, the WBS serves as a cost risk model of the program. Here, the summation of WBS element cost ranges across WBS levels forms a probability distribution of possible outcomes of total program cost, one of which is the point estimate. This is shown in Figure 19. The following applies a method of moments approach to the same WBS in Figure 15, which was used to illustrate the Monte Carlo simulation technique.

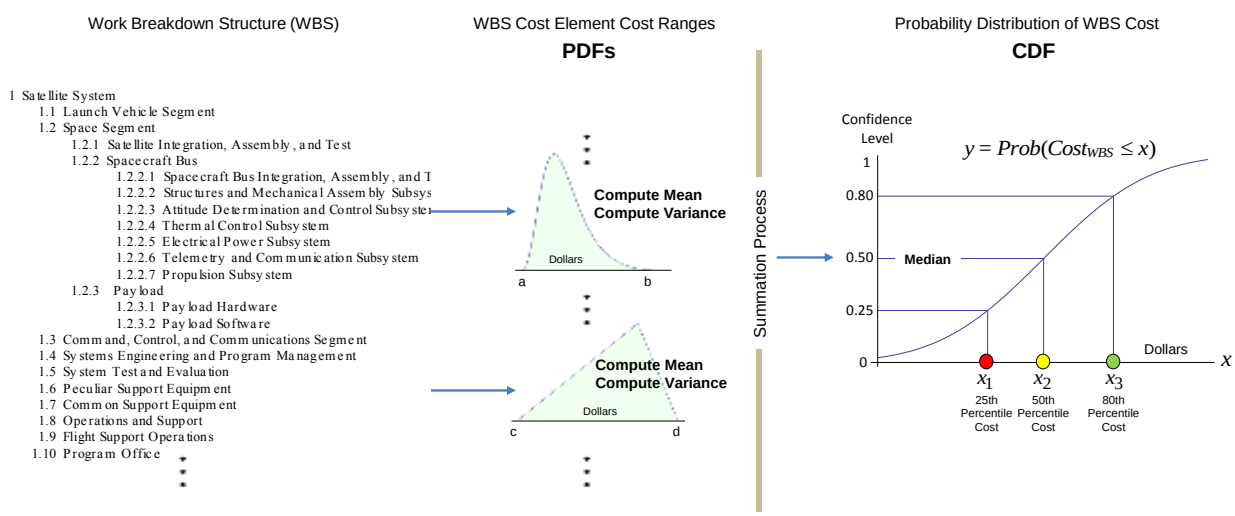


Figure 19. Method of Moments Applied to a WBS

Figure 20 presents a work breakdown structure consisting of the five cost elements X1, X2, X3, X4, and X5. A point estimate cost for each element is shown, along with an uncertainty distribution around each estimate. The cost mean and cost variance of each WBS element is analytically derived instead of empirically generated from random Monte Carlo samples. The cost means and cost variances of the WBS elements are summed and the probability distribution of $Cost_{WBS}$ is formed. In practice, this distribution is usually well approximated by a normal or a lognormal form[22]. Consider the following example.

[22] The following section will discuss the applicability of normal and lognormal forms in further detail.

Given Figure 20, define $Cost_{WBS}$ by Equation 11.

$$Cost_{WBS} = X1 + X2 + X3 + X4 + X5 \tag{11}$$

Thus, the mean and variance of $Cost_{WBS}$ is

$$E(Cost_{WBS}) = E(X1) + E(X2) + E(X3) + E(X4) + E(X5) \tag{12}$$

$$Var(Cost_{WBS}) = Var(X1) + Var(X2) + Var(X3) + Var(X4) + Var(X5) \tag{13}$$

From the information in Figure 20, it follows that

$$E(Cost_{WBS}) = 26.67 + 30 + 20 + 61.67 + 65 \tag{14}$$

$$E(Cost_{WBS}) = 203.34 \ (\$M)$$

$$Var(Cost_{WBS}) = 9.72 + 133.33 + 12.5 + 109.72 + 100 = 365.27 (\$M)^2 \tag{15}$$

$$\sigma_{Cost_{WBS}} = \sqrt{Var(Cost_{WBS})} = 19.11 (\$M) \tag{16}$$

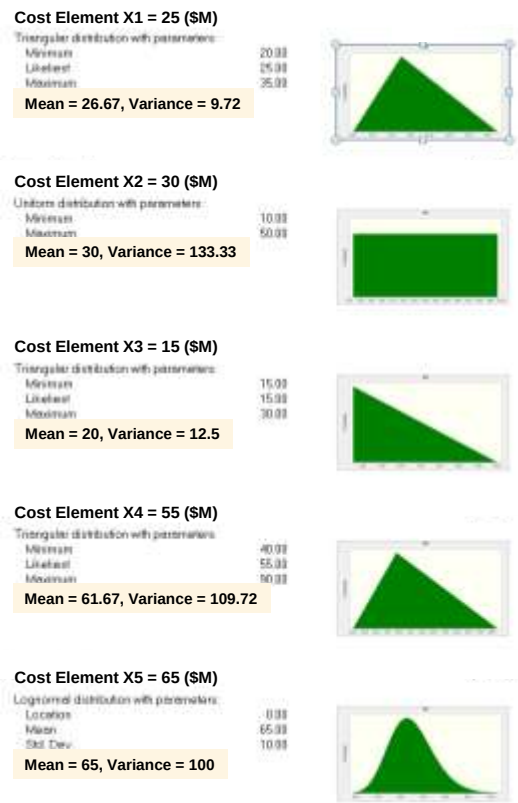


Figure 20. A WBS for Method of Moments

The mean and standard deviation computed by the method of moments, shown by Equations 14 and 16, are very close to these same statistics derived empirically by the Monte Carlo simulation summarized in Figure 16. Mutual independence between WBS element costs X1, X2, X3, X4, and X5 has been assumed in this example. Mutual independence implies these five WBS element costs are mutually uncorrelated. Thus, the variance of the sum of X1, X2, X3, X4, and X5 is the sum of their individual variances shown in Equation 15.

Figure 21 presents the probability distribution of $Cost_{WBS}$. The dots depict the probability distribution of $Cost_{WBS}$ generated by a Monte Carlo simulation of the same WBS in Figure 15. The simulation produced a mean of $Cost_{WBS}$ equal to $203.3M and a standard deviation of $Cost_{WBS}$ equal to $19M. In Figure 21, the solid red line is the probability distribution of $Cost_{WBS}$ assuming its possible cost outcomes fall along a normal probability distribution – with mean and variance derived by the method of moments. The solid blue line is the probability distribution of $Cost_{WBS}$ assuming its possible cost outcomes fall along a lognormal probability distribution – with mean and variance derived by the method of moments.

In Figure 21, the simulated, normal, and lognormal probability distributions are almost indistinguishable. Mentioned earlier, a reason for this is the assumed mutual independence between WBS element costs X1, X2, X3, X4, and X5 in this example. The Central Limit Theorem (CLT) enters the picture and ensures the eventual tendency of the simulated distribution to approach a normal distribution. Figure 21 visually reveals the lognormal distribution also approximates the normal or simulated distributions of $Cost_{WBS}$. The goodness of these approximations is often seen in cost risk analysis. Reasons and factors for this are further discussed in Section 3.2.2. Appendix B offers a further illustration of the method of moments applied to a larger work breakdown structure than above, with correlated WBS element costs.

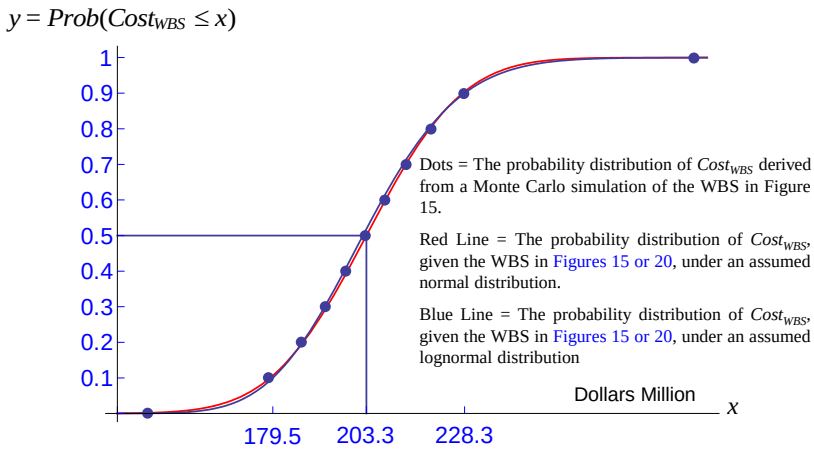


Figure 21. Probability Distribution of $Cost_{WBS}$

3.2.2 Method of Moments: Forms of the Probability Distribution of Total Cost

The preceding illustrated the method of moments applied to a WBS consisting of five cost elements. This section extends the method of moments to a work breakdown structure of *n* cost elements. Let

$$Cost_{WBS} = X_1 + X_2 + X_3 + \ldots + X_n \tag{17}$$

It then follows that the mean or expected value of $Cost_{WBS}$ is

$$E(Cost_{WBS}) = E(X_1) + E(X_2) + E(X_3) + \ldots + E(X_n) \tag{18}$$

If the costs of cost elements X_1, X_2, X_3, …, X_n are independent, then the variance of $Cost_{WBS}$ is

$$Var(Cost_{WBS}) = Var(X_1) + Var(X_2) + Var(X_3) + \ldots + Var(X_n) \tag{19}$$

If the costs of cost elements X_1, X_2, X_3, …, X_n are **not** independent, then the variance of $Cost_{WBS}$ is

$$\begin{aligned} Var(Cost_{WBS}) &= Var(X_1) + Var(X_2) + Var(X_3) + \ldots + Var(X_n) \\ &\quad + 2\rho_{X_1,X_2}\sigma_{X_1}\sigma_{X_2} + 2\rho_{X_1,X_3}\sigma_{X_1}\sigma_{X_3} + \ldots + 2\rho_{X_i,X_j}\sigma_{X_i}\sigma_{X_j} \end{aligned} \tag{20}$$

for all *i* and *j* such that $1 \le i \le j \le n$. In Equation 20, ρ_{X_i,X_j} is the Pearson product-moment correlation between the costs of WBS element random variables X_i and X_j (discussed in the previous section).

Once values for $E(Cost_{WBS})$ and $Var(Cost_{WBS})$ are computed they are used in the method of moments to specify a lognormal or normal probability distribution of $Cost_{WBS}$, with mean $E(Cost_{WBS})$ and variance $Var(Cost_{WBS})$. The following illustrates how to create these probability distributions in Excel.

How to Specify a Lognormal Probability Distribution

In the discussion illustrated in Figure 20, the method of moments was used to derive the mean and variance of $Cost_{WBS}$; specifically, it was shown that

$$E(Cost_{WBS}) = 203.34 \text{ (\$M) and } Var(Cost_{WBS}) = 365.27 \text{ (\$M)}^2$$

Microsoft Excel can be used to specify a lognormal distribution with these computed statistics. This is illustrated by the following steps.

Step 1. Transform $E(Cost_{WBS})$ and $Var(Cost_{WBS})$
Transform the mean and variance of $Cost_{WBS}$, computed by the method of moments, into the lognormal parameters μ and σ^2 given by Equations 21 and 22

$$\mu = \frac{1}{2}\ln\left[\frac{a^4}{a^2+b}\right] \tag{21}$$

$$\sigma^2 = \ln\left[\frac{a^2+b}{a^2}\right] \tag{22}$$

where $a = E(Cost_{WBS})$ and $b = Var(Cost_{WBS})$.

Step 2. Compute and Enter Parameters μ and σ^2 into Excel
In this step, we compute the values for μ and σ^2 given by Equations 21 and 22. Define

$$a = E(Cost_{WBS}) = 203.34 \text{ (\$M)}$$

$$b = Var(Cost_{WBS}) = 365.27 \text{ (\$M)}^2$$

From Equations 21 and 22, it follows that

$$\mu = \frac{1}{2}\ln\left[\frac{a^4}{a^2+b}\right] = \frac{1}{2}\ln\left[\frac{(203.34)^4}{(203.34)^2+365.27}\right] = 5.31048 \tag{23}$$

$$\sigma^2 = \ln\left[\frac{a^2+b}{a^2}\right] = \ln\left[\frac{(203.34)^2+365.27}{(203.34)^2}\right] = 0.00879543 \tag{24}$$

$$\sigma = \sqrt{\sigma^2} = 0.093784 \tag{25}$$

From the above, the values for a, b, μ, and σ are used to form the lognormal probability distribution of $Cost_{WBS}$. Table 4 is one way to use Excel with outputs from the method of moments to derive the lognormal probability distribution of $Cost_{WBS}$. Columns A, B, and C are inputs from Step 2. Column E is the probability α that $Cost_{WBS}$ does not exceed x dollars; that is $Prob(Cost_{WBS} \le x) = \alpha$. In Table 4, observe how the values computed by Excel in Column F track to the lognormal probability distribution shown by the blue line in Figure 21.

How to Specify a Normal Probability Distribution
Excel can be used to specify a normal probability distribution of $Cost_{WBS}$, with mean $E(Cost_{WBS})$ and $Var(Cost_{WBS})$ derived from the method of moments. Table 5 illustrates how to create this distribution in Excel. Columns A, B, and C are the computed values of $E(Cost_{WBS})$ and $Var(Cost_{WBS})$ derived from the method of moments. Column E is the probability α that $Cost_{WBS}$ does not exceed x dollars; that is $Prob(Cost_{WBS} \le x) = \alpha$. In Table 5, observe how the values computed by Excel in Column F track to the normal probability distribution shown by the red line in Figure 21.

	A	B	C	D	E	F
1					IF	THEN
2	EQUATION				α =	x =
3	12	a =	203.34		0.00	151.5
4	14	b =	365.27		0.10	179.5
5	17	μ =	5.31048		0.20	187.1
6	19	σ =	0.093784		0.30	192.7
					0.40	197.7
	Cumulative Lognormal Probability Distribution of $Cost_{WBS}$				0.50	202.4
					0.60	207.3
					0.70	212.7
					0.80	219.1
					0.90	228.3
					1.00	270.5

F3 = LOGNORM.INV(E3,C5,C6)
F4 = LOGNORM.INV(E4,C5,C6)
F5 = LOGNORM.INV(E5,C5,C6)
F6 = LOGNORM.INV(E6,C5,C6)

Table 4. Excel Function to Compute Lognormal Probabilities for $Cost_{WBS}$

	A	B	C	D	E	F
1					IF	THEN
2	EQUATION				α =	x =
3	12	a =	203.34		0.00	144.3
4	14	b =	365.27		0.10	178.8
5		$\sqrt{b}$ =	19.11		0.20	187.3
6					0.30	193.3
					0.40	198.5
	Cumulative Normal Probability Distribution of $Cost_{WBS}$				0.50	203.3
					0.60	208.2
					0.70	213.4
					0.80	219.4
					0.90	227.8
					1.00	262.4

F3 = NORM.INV(E3,C5,C6)
F4 = NORM.INV(E4,C5,C6)
F5 = NORM.INV(E5,C5,C6)
F6 = NORM.INV(E6,C5,C6)

Table 5. Excel Function to Compute Normal Probabilities for $Cost_{WBS}$

Practice Point 13: Method of Moments Approximation of the Distribution of Cost

This practice point provides guidance for approximating the probability distribution of a program's total cost $Cost_{WBS}$. Seen in the preceding discussion, the method of moments produces an analytically derived measure of the mean and variance of $Cost_{WBS}$. Monte Carlo simulation produces an empirically derived basis for these two measures, as well as an empirically derived probability distribution of $Cost_{WBS}$. To produce the probability distribution of $Cost_{WBS}$ using the method of moments, the form or shape of this distribution must be assumed. Best practice observations, published evidence, and statistical tests indicate the probability distribution of $Cost_{WBS}$ is often well approximated by normal or lognormal forms. This is

seen in Figure 21. In Figure 21, the normal and lognormal distributions well approximate the empirically derived probability distribution of $Cost_{WBS}$ – shown by the dots generated by the Monte Carlo simulation. There are many technical and empirically observed reasons for this. Primary among them is that a program's total cost is a summation of WBS element costs, including a summation of costs derived from nonlinear cost estimation relationships. Within the WBS, it is typical to have a mixture of independent and correlated element costs. The greater the number of independent WBS element costs, the more it is that the probability distribution of $Cost_{WBS}$ is approximately normal. Why is this? Mentioned earlier, it is essentially the phenomenon explained by the Central Limit Theorem.

The Central Limit Theorem is very powerful in that it does not take many independent WBS element costs for the probability distribution of $Cost_{WBS}$ to approach normality. This central tendency is evidenced when (1) a sufficient number of independent WBS element costs are summed and (2) no WBS element's probability distribution has a much larger standard deviation than the standard deviations of the other WBS element distributions. When conditions in the WBS result in $Cost_{WBS}$ being positively skewed (i.e., a non-normal distribution), then the lognormal often well approximates the distribution function of $Cost_{WBS}$. There is an extensive theoretical and practical discussion on this topic in Garvey (2000) and Young (1995), as well from other published research by the cost analysis community.

3.2.3 Method of Moments: A Scenario-Based Implementation

This section presents a newly published variation on the method of moments called the enhanced Scenario-Based Method (eSBM)[23]. eSBM is a historical data driven approach for cost risk analysis. Historical cost risk data directly integrates into the eSBM algorithms to produce a range of possible costs and measures of cost estimate confidence, driven by past program histories.

eSBM is a method of moments alternative to Monte Carlo simulation. With its simplified analytics, eSBM eases the mathematical burden on analysts, focusing instead on defining and analyzing risk scenarios as the basis for deliberations on the amount of cost reserve needed to protect a program from unwanted or unexpected cost increases. With today's emphasis on affordability-based decision-making, eSBM promotes realism in estimating costs by providing an analytically traceable and defensible basis behind historical data-derived measures of risk and cost estimate confidence.

eSBM and Affordability-Based Decision Making

Systems engineering is more than developing and employing inventive technologies. Designs must be adaptable to change, evolving demands of users, and resource constraints. They must be balanced with respect to performance and affordability goals while being continuously risk managed throughout a system's life cycle. Systems engineers and managers must also understand the social, political, and economic environments within which the system operates. These factors can significantly influence affordability, design trades, and resultant investment decisions.

In the Department of Defense, affordability means conducting a program at a cost constrained by the maximum resources the Department can allocate for that capability. Affordability is the lever that constrains system designs and requirements. With the Department implementing affordability-based decision making at major milestones, identifying affordability risk drivers requires a rigorous assessment

[23] eSBM has been published in the following US government cost risk analysis guides:

- United States Air Force, Air Force Cost Analysis Agency, 2007. *Cost Risk and Uncertainty Analysis Handbook (CRUH).*
- NASA 2008. *Cost Estimating Handbook.*
- GAO 2009. *Cost Estimating and Assessment Guide*, GAO-09-3SP.
- Missile Defense Agency (MDA), 2012. *Cost Estimating and Analysis Handbook*.
- United States Air Force, Air Force Cost Analysis Agency, 2013. *Joint Cost-Schedule Risk and Uncertainty Handbook*.

and quantification of cost risk. With this, the trade space around these drivers can be examined for opportunities to eliminate or manage affordability threats before they materialize.

Pressures on acquisition programs to deliver systems that meet cost, schedule, and performance are omnipresent. Illustrated in Figure 22, risk becomes an increasing reality when stakeholder expectations push what is technically or economically feasible. Managing risk is managing the inherent contention that exists within and across these dimensions.

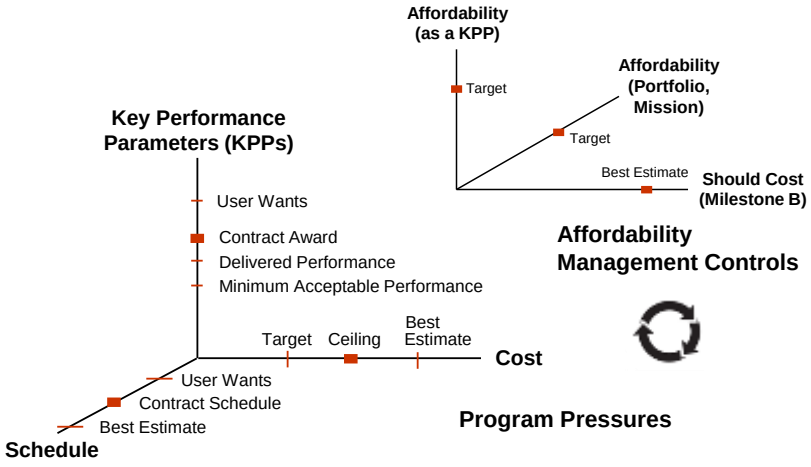


Figure 22. Program Pressures and Affordability Management Controls

Recognizing this, the Department of Defense has instituted management controls to maintain the affordability of programs and the capability portfolios where many programs reside. Shown in Figure 22, affordability is now a key performance parameter (KPP) with its target set as a basis for pre-milestone B (MS B) decisions and engineering tradeoff analysis [USD (AT&L) Memorandum, 3 November 2010].

Managing to affordability must consider the potential consequences of risks to programs and their portfolios, particularly during pre-MS B design trades. When a new program is advocated for a portfolio, or mission area, a cost risk analysis can derive measures of confidence in the adjustments needed to absorb the program. For milestone B decisions, risk-adjusted cost tradeoff curves can be developed to identify and manage affordability driving risk events that threaten a program's integrity and life cycle sustainability.

Shown in Figure 22, a management control called "should cost" is now exercised following a milestone B decision. "Should cost" is an approach to life cycle cost management that is focused on finding ways a program can be delivered below its affordability target. Achieving this means successfully managing risk and its cost impacts, as they are quantified in the program cost estimate or its independent cost estimate.

The enhanced scenario-based method was developed in two forms, the non-statistical eSBM and the statistical eSBM, the latter of which is the form needed for generating a probability distribution of a program's total cost. The following discussion describes each eSBM form and their mutual relationship.

The Non-Statistical eSBM

The enhanced scenario-based method is centered on articulating and costing a program's risk scenarios. Risk scenarios are coherent stories or narratives about potential events that, if they occur, increase program cost beyond what was planned.

Practice Point 14: The process of defining risk scenarios or narratives is a good practice. It builds the rationale and case arguments to justify the reserve needed to protect program cost from the realization of unwanted events. This is lacking in Monte Carlo simulation if designed as arbitrary randomizations of possible program costs, a practice which can lead to reserve recommendations absent clear program context for what these funds are to protect. The following presents the non-statistical eSBM.

Figure 23 illustrates the process flow of the non-statistical implementation of eSBM. The first step is input to the process. It is the program's point estimate cost (PE). For this paper, the point estimate cost is the cost that does not include allowances for uncertainty. The PE is the sum of the WBS element costs across the program's work breakdown structure *without adjustments for uncertainty*. The PE is often developed from the program's cost analysis requirements description document (CARD).

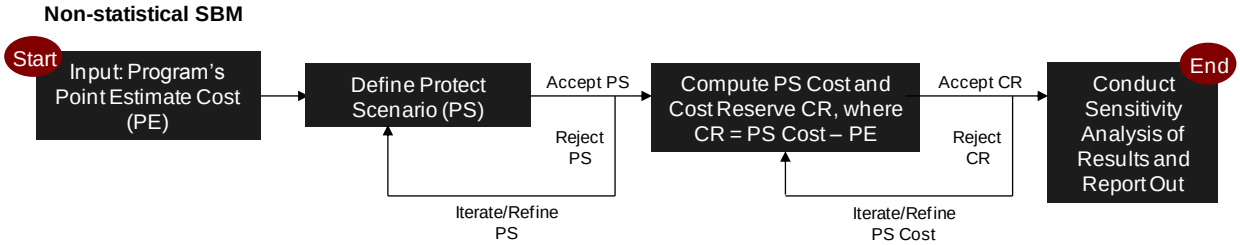


Figure 23. The Non-statistical eSBM Process

The next step in Figure 23 is defining a protect scenario. A protect scenario captures the cost impacts of major known risks to the program – those events the program must monitor and guard against occurring. The protect scenario is not arbitrary, nor should it reflect extreme worst-case events. It should reflect a possible program cost that, in the judgment of the program, has *an acceptable chance of not being exceeded*. In practice, it is envisioned that management will converge on an "official" protect scenario after deliberations on the one initially defined. This part of the process ensures all parties reach a consensus understanding of the program's risks and how they are best described by the protect scenario.

Once the protect scenario is established its cost is then estimated. Denote this cost by PS. The amount of cost reserve dollars (CR) needed to protect program cost can be computed as the difference between the PS and the PE. Shown in Figure 23, there may be additional refinements to the cost estimated for the protect scenario, based on management reviews and other considerations. The process may be iterated until the reasonableness of the magnitude of the cost reserve dollars is accepted by management.

The final step in Figure 23 is a sensitivity analysis to identify critical drivers associated with the protect scenario and the program's point estimate cost. It is recommended that the sensitivity of the amount of reserve dollars, computed in the preceding step, be assessed with respect to variations in the parameters associated with these drivers.

The non-statistical eSBM, though simple in appearance, is a form of cost risk analysis. The process of defining risk scenarios is a valuable exercise in identifying technical and cost estimation challenges inherent to the program. Without the need to define risk scenarios, cost risk analyses can be superficial, its case-basis not defined or carefully thought through. Scenario definition encourages a discourse on risks that otherwise might not be held, thereby allowing risks to become fully visible, traceable, and estimative to program managers and decision-makers.

The non-statistical eSBM, in accordance with its non-statistical nature, does not produce confidence measures. The chance that the protect scenario cost, or of any other defined risk scenario's cost, will not

be exceeded is not explicitly determined. The question is "*Can this eSBM implementation be modified to produce confidence measures while maintaining its simplicity and analytical features?*" The answer is yes, and a way to approach this excursion is presented next.

Statistical eSBM

This section presents a statistical eSBM. Instead of a Monte Carlo simulation, the statistical eSBM is a closed-form method of moments approach, requiring only a look-up table and a few equations.

Among the many reasons to implement a statistical track in eSBM are the following: (1) it enables affordability-level confidence measures to be determined, (2) it offers a way for management to examine changes in confidence measures as a function of how much reserve to buy to increase the chance of program success, and (3) it provides an ability to measure where the protect scenario cost falls on the probability distribution of the program's total cost.

Figure 24 illustrates the process flow of the statistical eSBM. The upper part replicates the process steps of the non-statistical eSBM, and the lower part appends the statistical eSBM process steps. Thus, the statistical eSBM is an augmentation of the non-statistical eSBM.

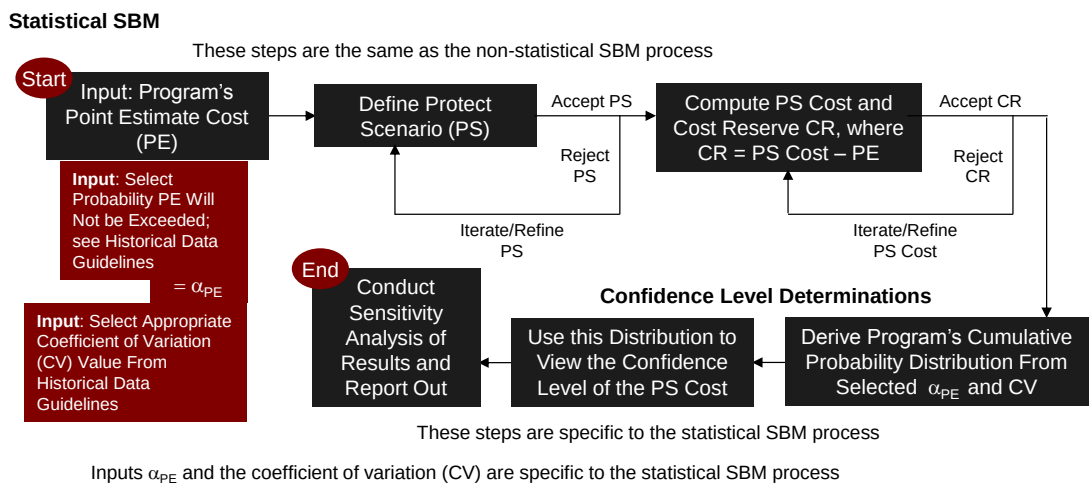


Figure 24. The Statistical eSBM Process

To apply the statistical eSBM, three inputs shown on the left in Figure 24 are required. These are the PE, the probability that PE will not be exceeded, and the coefficient of variation (CV), which is explained below. The PE is the same as previously defined in the non-statistical eSBM. The probability that PE will not be exceeded is the value α, such that

$$P(Cost_{WBS} \le PE) = \alpha \tag{26}$$

In Equation 26, $Cost_{WBS}$ is the true but uncertain total cost of the program and PE is the program's point estimate cost. The probability α is a judged value guided by historical experience that it typically falls in the interval $0.10 \le \alpha \le 0.50$. This interval reflects the understanding that a program's point estimate cost PE usually faces higher, not lower, probabilities of being exceeded.

The coefficient of variation (CV) is the ratio of a probability distribution's standard deviation to its mean. This ratio is given by Equation 27. The CV is a way to examine the variability of any distribution at plus or minus one standard deviation around its mean.

$$CV = \frac{\sigma}{\mu} \tag{27}$$

With values assessed for α and CV, the cumulative probability distribution of $Cost_{WBS}$ can then be derived. This distribution is used to view the confidence level associated with the protect scenario cost PS, as well as confidence levels associated with any other cost outcome along this distribution.

The final step in Figure 24 is a sensitivity analysis. Here, we can examine the kinds of sensitivities previously described in the non-statistical SBM implementation, as well as uncertainties in values for α and CV. This allows a broad assessment of confidence level variability, which includes determining a range of possible program cost outcomes for any specified confidence level.

Figure 25 illustrates an output from the statistical eSBM process. This is a normal probability distribution with point estimate cost PE equal to \$100M, α set to 0.25, and CV set to 0.50. The range \$75M to \$226M is plus or minus one standard deviation σ around the mean of \$151M.

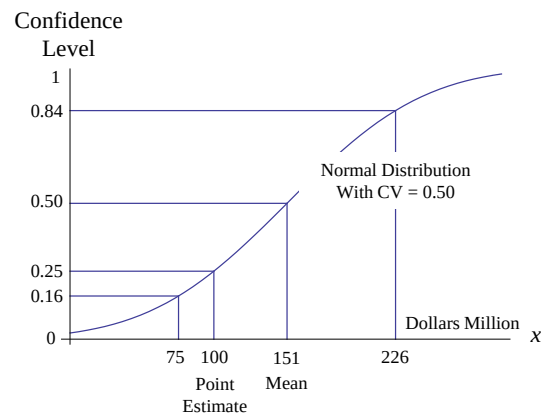


Figure 25. Statistical eSBM Confidence Intervals

Statistical eSBM Equations: For Normal Distribution of $Cost_{WBS}$

The following equations derive from the assumption that a program's total cost, denoted by $Cost_{WBS}$, is normally distributed and the point (PE, α) falls along this distribution. Given PE, α, and CV, then the mean and standard deviation of $Cost_{WBS}$ are given by the following:

$$\mu = PE - z\frac{(CV)PE}{1 + z(CV)} \tag{28}$$

$$\sigma = \frac{(CV)PE}{1 + z(CV)} \tag{29}$$

where CV is the coefficient of variation, PE is the program's point estimate cost, and z is the value such that $P(Z \le z) = \alpha$ where Z is the standard normal random variable. Values for z are available in look-up tables for the standard normal [Garvey, 2000] or from the Excel function z = *Norm.S.Inv*(*percentile*); e.g., z = 0.525 = *Norm.S.Inv*(0.70). With the values computed from Equations 28 and 29, the normal distribution function of $Cost_{WBS}$ is fully specified, along with the probability that $Cost_{WBS}$ may take any particular outcome, such as the protect scenario cost PS.

Statistical eSBM Equations: For Lognormal Distribution of $Cost_{WBS}$

The following equations derive from the assumption that a program's total cost, denoted by $Cost_{WBS}$, is lognormally distributed and the point (PE, α) falls along this distribution. Given PE, α, and CV, then the mean and standard deviation of $Cost_{WBS}$ are given by the following:

$$\mu = e^{a + \frac{1}{2}b^2} \tag{30}$$

$$\sigma = \sqrt{e^{2a+b^2}(e^{b^2} - 1)} = \mu\sqrt{(e^{b^2} - 1)} \tag{31}$$

where

$$a = \ln PE - z\sqrt{\ln(1+(CV)^2)} \tag{32}$$

$$b = \sqrt{\ln(1+(CV)^2)} \tag{33}$$

With the values computed from Equations 30 and 31, the lognormal distribution function of $Cost_{WBS}$ is fully specified, along with the probability that $Cost_{WBS}$ may take any particular outcome, such as the protect scenario cost PS. Appendix C presents a numerical example that illustrates implementing the above eSBM lognormal equations.

Example 1 (Normal Distribution)
Suppose the distribution function of a program's total cost is normal. Suppose the program's point estimate cost is $100M and this was assessed to fall at the 25th percentile. Suppose the type and life cycle phase of the program is such that 30 percent variability in cost around the mean has been historically seen. Suppose the protect scenario was defined and determined to cost $145M.

A) Compute the mean and standard deviation of $Cost_{WBS}$.
B) Plot the distribution function of $Cost_{WBS}$.
C) Determine the confidence level of the protect scenario cost and its associated cost reserve.
D) Determine the program cost outcome at the 80th percentile confidence level, denoted by $x_{0.80}$.

Solution
A) From Equations 28 and 29

$$\mu = PE - z\frac{(CV)PE}{1+z(CV)} = 100 - z\frac{(0.30)(100)}{1+z(0.30)}$$

$$\sigma = \frac{(CV)PE}{1+z(CV)} = \frac{(0.30)(100)}{1+z(0.30)}$$

We need z to complete these computations. Since the distribution function of $Cost_{WBS}$ was given to be normal, it follows that $P(Cost_{WBS} \le PE) = \alpha = P(Z \le z)$, where Z is a standard normal random variable. Values for z are available in Excel and are computed as follows. Given $\alpha = 0.25$ in this example, then enter this formula into Excel: $\text{NORM.S.INV}(0.25)$; that is,

$$z = \text{NORM.S.INV}(\alpha) = \text{NORM.S.INV}(0.25) = -0.6745$$

Therefore,

$$\mu = PE - z\frac{(CV)PE}{1+z(CV)} = 100 - (-0.6745)\frac{(0.30)(100)}{1+(-0.6745)(0.30)} = 125.4\,(\$\text{M})$$

$$\sigma = \frac{(CV)PE}{1+z(CV)} = \frac{(0.30)(100)}{1+(-0.6745)(0.30)} = 37.6\,(\$\text{M})$$

B) A plot of the probability distribution function of $Cost_{WBS}$ is shown in Figure 26. This is a normal distribution with mean $125.4M and standard deviation $37.6M, as determined from Part A).

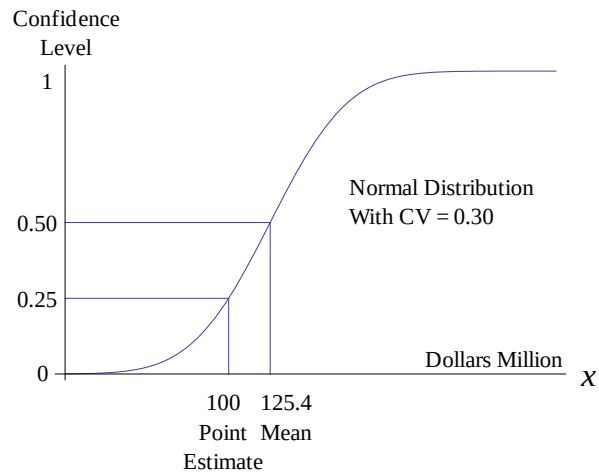


Figure 26. Normal Distribution Function of $Cost_{WBS}$

C) To determine the confidence level of the protect scenario, find α_{PS} such that

$$P(Cost_{WBS} \le PS = 145) = \alpha_{PS}$$

Finding α_{PS} is equivalent to solving the expression $\mu + z_{PS}\sigma = PS$ for z_{PS}. From this,

$$z_{PS} = \frac{PS - \mu}{\sigma} = \frac{PS}{\sigma} - \frac{1}{CV}$$

Since $PS = 145$, $\mu = 125.4$, and $\sigma = 37.6$ it follows that

$$z_{PS} = \frac{PS - \mu}{\sigma} = \frac{PS}{\sigma} - \frac{1}{CV} = \frac{145}{37.6} - \frac{1}{(0.30)} = 0.523$$

Thus, we want α such that $P(Z \le z_{PS} = 0.523) = \alpha$. Values for α are available in Excel as follows. With $z_{PS} = 0.523$, enter into Excel: NORM.S.DIST(0.523,TRUE); that is,

$$\alpha = \text{NORM.S.DIST}\left(z_{PS}, \text{TRUE}\right) = \text{NORM.S.DIST}\left(0.523, \text{TRUE}\right) = 0.70$$

Therefore, the $145M protect scenario cost falls at the 70th percentile of the distribution. This implies a cost reserve CR equal to $45M.

D) To determine the 80th percentile confidence level, we need to find $z_{0.80}$ such that

$$P(Z \le z_{0.80}) = 0.80$$

Given $\alpha = 0.80$ in this example, enter into Excel: $\text{NORM.S.INV}(0.80)$; that is,

$$z_\alpha = \text{NORM.S.INV}(\alpha) = z_{0.80} = \text{NORM.S.INV}(0.80) = 0.8416$$

Substituting $\mu = 125.4$ and $\sigma = 37.6$ (determined in Part A) yields the following:

$$\mu + z_{0.80}\sigma = 125.4 + 0.8416(37.6) = x_{0.80} = 157$$

Therefore, the cost associated with the 80th percentile confidence level is $157M. Figure 27 presents a summary of the results in this example.

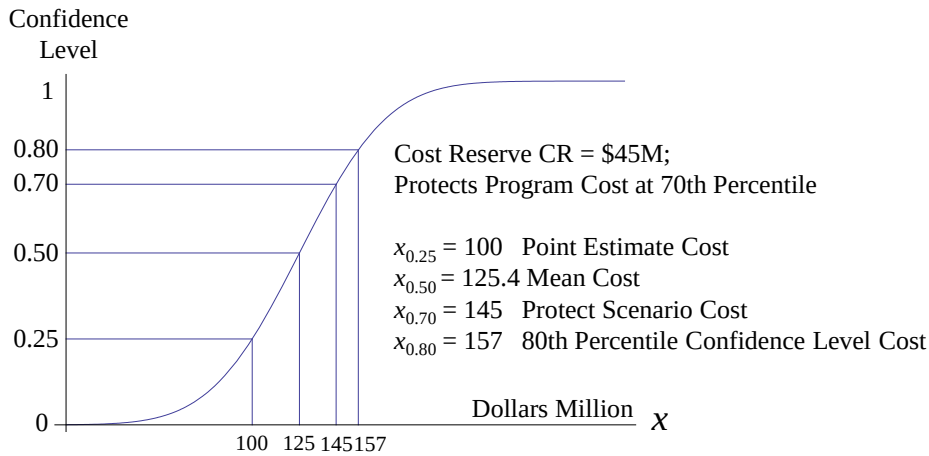


Figure 27. Normal Distribution Function of $Cost_{WBS}$ for Various Confidence Intervals

eSBM Sensitivity Analysis

This section shows how eSBM can assess the sensitivity in program cost $Cost_{WBS}$ at the 80th percentile, or any chosen confidence level along a probability distribution of program cost. Example 1 is used to illustrate these ideas.

In Example 1, single values for α and CV were used. If a range of possible values is used, then a range of possible program costs can be generated at any percentile along the distribution. For instance, suppose historical cost data for a particular program indicates its CV varies in the interval $0.20 \le CV \le 0.50$. Given the conditions in Example 1, variability in CV affects the mean and standard deviation of program cost. This is illustrated in Table 6, given a program's point estimate cost equal to $100M and its $\alpha = 0.25$.

Coefficient of Variation (CV)	Standard Deviation ($M)	Mean ($M) 50th percentile*	Confidence Level ($M) 80th percentile
0.20	23.1	115	135
0.30	37.6	125	157
0.40	54.8	137	183
0.50	75.4	151	214

Table 6. Ranges of Cost Outcomes in Confidence Levels

Table 6 shows a range of possible cost outcomes for the 50th and 80th percentiles. Selecting a particular outcome can be guided by the CV considered most representative of the program's uncertainty at its specific life cycle phase. This is guided by the scenario or scenarios developed at the start of the eSBM process. Figure 28 graphically illustrates the results in Table 6.

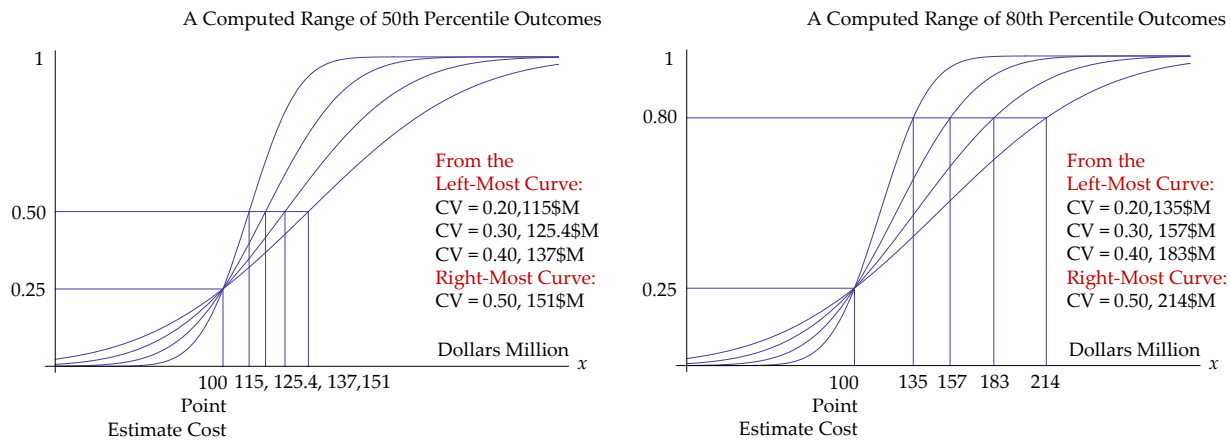


Figure 28. A Range of Confidence Level Cost Outcomes

eSBM Inputs

Two inputs are needed for the statistical eSBM. They are (1) the probability α that a program's point estimate cost will not be exceeded and (2) the program's coefficient of variation CV. With just these inputs, statistical measures of cost risk and confidence are produced. The following presents ways to assess α and CV, with the use of historical data as a guide.

Discussed earlier, the probability a program's point estimate cost PE will not be exceeded is the value α such that $P(Cost \le PE) = \alpha$. Anecdotally it is well understood a program's PE usually faces higher, not lower, probabilities of being exceeded – especially in the early life cycle phases. The interval $0.10 \le \alpha \le 0.50$ expresses this anecdotal experience. It implies a program's PE will very probably experience growth instead of reduction. Unless there are special circumstances, a value for α from this interval should be selected for eSBM and a justification written for the choice. Yet, the question remains what value of α should be chosen?

To address this question, the Naval Center for Cost Analysis (NCCA) recently developed a large dataset of historical CVs from hundreds of major defense programs[24]. From this, we now have ways to guide the choice α and CV from their analysis of program cost growth histories published in articles by Garvey (2012) and Braxton, Flynn, and Lee, (2012). For example, the NCCA data revealed a historical CV for a set of Milestone B Navy programs as

$$CV = 0.51 = \frac{0.69}{1.36} = \frac{\sigma}{\mu}$$

If this CV follows a lognormal distribution[25] (with $\sigma = 0.69$ and $\mu = 1.36$) then it can be shown that the mean cost growth factor falls at the 59th percentile confidence level. This is shown in Figure 29.

A program's point estimate cost PE is the baseline from which cost growth is applied. Thus, PE has a cost growth factor (CGF) equal to one. In Figure 29, this is shown by x = 1. For the historical programs with

[24] The collection and analysis of historical data for use in eSBM was under the auspices of Ms. Wendy Kunc, (Deputy Assistant Secretary for Cost and Economics, Office of the Assistant Secretary of the Navy (Financial Management & Comptroller) and Executive Director of the Naval Center for Cost Analysis (NCCA). The analysts who led this aspect of eSBM were Dr. Brian Flynn, Mr. Peter Braxton, and Mr. Richard Lee of Technomics, Inc.

[25] Findings from a RAND study on weapon system cost growth [RAND, 2007] revealed the distribution of cost growth is approximately lognormal.

CV represented by the lognormal distribution in Figure 29, it can be shown that x = 1 falls at the 34th percentile confidence level. This means $\alpha = 0.34$ for these program histories and we can write

$$CV = 0.51 = \frac{0.69}{1.36} = \frac{\sigma}{\mu} \Rightarrow \alpha = 0.34$$

This discussion shows how an empirical α can be derived from program cost growth histories to guide the choice of its value in eSBM. The historical cost growth data developed by the NCCA enables deriving insights into point estimate confidence by major acquisition milestone. Shown above, an analysis of Milestone B program histories indicate a probability of 0.34 that program point estimates will not be exceeded. In eSBM, this is the value α such that $P(Cost \le PE) = \alpha$. This is the first time an historical, data-driven, insight into point estimate confidence has been derived. It furthers the otherwise anecdotal experience that α often falls in the interval $0.10 \le \alpha \le 0.50$ for programs in these life cycle phases.

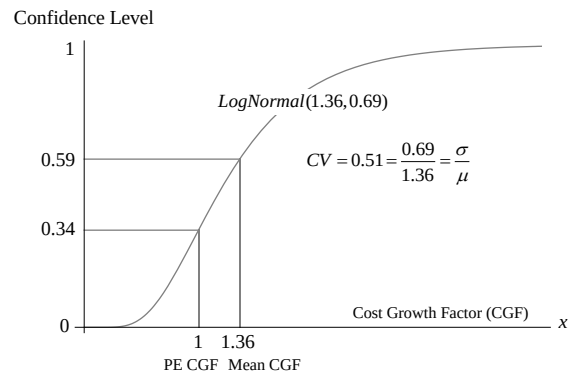


Figure 29. Deriving α From Historical Program CVs

Table 7 presents a summary of the NCCA of historical CVs for estimates of total program acquisition cost. These data can be used to guide the choice of values for CVs in eSBM. Further information on the data in Table 7 is offered in Garvey, Flynn, Braxton, and Lee (2012).

CVs for Estimates of Total Program Acquisition Cost
(Based on Three Samples: All Data; Data => 1980s; and Data =>1990s)

CV Calculation and Definition	Milestone A			Milestone B			Milestone C		
	All	=>1980	=>1990	All	=>1980	=>1990	All	=>1980	=>1990
Quantity Random									
Then-year dollars	1.20	0.84	0.77	0.87	0.69	0.53	0.53	0.53	0.29
Base-year dollars	0.82	0.85	0.76	0.63	0.65	0.51	0.45	0.45	0.25
Quantity Exogenous	Max		Min	Max		Min	Max		Min
Then-year dollars	0.76	0.54	0.50	0.51	0.40	0.33	0.26	0.26	0.16
Base-year dollars	0.52	0.56	0.50	0.36	0.37	0.30	0.19	0.19	0.11

Table 7. CVs for Estimates of Total Program Acquisition Cost

Summary

eSBM was developed as an alternative to advanced statistical methods for generating measures of cost risk. With today's emphasis on affordability-based decision-making, eSBM fosters realism in estimating costs by providing an analytically traceable and defensible basis behind data-derived measures of risk and cost estimate confidence.

Features of eSBM include the following:

- The ability to establish a range of possible program costs and measures of cost estimate confidence, driven by historical program performance;

- A focus on defining and analyzing risk scenarios as the basis for deliberations on the amount of cost reserve needed to protect a program from unwanted or unexpected cost increases;
- Provides an analytic argument for deriving the amount of cost reserve needed to guard against well-defined scenarios;
- Brings the discussion of scenarios and their credibility to decision-makers; this is a more meaningful topic to focus on instead of statistical abstractions cost risk analysis approaches can create;
- The ability to generate measures of cost estimate confidence in a spreadsheet environment;
- Does not require analysts to develop probability distribution functions for all the uncertain variables in a WBS, which can be time-consuming and hard to justify;
- Correlation is captured implicitly in the analysis by the chosen coefficient of variation – thereby eliminating the need to address this measure explicitly.
- The approach fully supports traceability and focuses attention on key program risk events, identified in the written scenarios that have the potential to drive cost higher than expected.

Practice Point 15: Cost risk analysis inputs should have a traceable and defensible basis in terms of their origin, pedigree, and soundness. Inputs derived from or based on evidence, historical data, or subject opinion should be documented and summarized in ways that support independent reviews.

Practice Point 16: As a management practice, encourage and emphasize a careful and deliberative approach to cost risk analysis, regardless of the analytical method employed. Require the development of realistic excursions from a system's technical baseline or its cost analysis requirements description document (CARD) that represent its risk scenarios.

Analyze the consequences of identified risk scenarios on cost. Use these findings as a basis for identifying the amount of cost reserve needed to protect the budget from unexpected cost increases. Read from the cost probability distribution the percentile associated with the recommended cost reserve, to determine its level of confidence instead of arbitrarily budgeting to a predetermined confidence level. Time is best spent building the analysis and case arguments to justify how a confluence of identified risk events, that form one or more risk scenarios, may drive the cost of a program to a particular percentile. This is the perspective from which to make risk-informed budget and cost risk reserve decisions.

4.0 SUMMARY

The following provides a set of considerations and recommended practices when performing cost risk analyses. A summary of historical cost risk data and its supporting data sources are presented.

Practice Point 17: Treating Cost as a Random Variable — The cost of a future system can be significantly affected by uncertainty. The existence of uncertainty implies the existence of a range of possible costs. How can a decision-maker be shown the chance a particular cost in the range of possible costs will be realized? The probability distribution is a recommended approach for providing this insight.

Probability distributions result when independent variables (e.g., weight, power-output, schedule) used to derive a system's cost randomly assume values across ranges of possible values. For instance, the cost of a satellite might be derived using a range of possible weight values, with each value randomly occurring. This approach treats cost as a random variable. It is recognition that values for these variables (such as weight) are not typically known with sufficient precision to predict cost perfectly, *at a time when such predictions are needed*. This point is further articulated by the well-respected analyst S. A. Book[26].

> "*The mathematical vehicle for working with a range of possible costs is the probability distribution, with cost itself viewed as a "random variable". Such terminology does not imply, of course, that costs are "random" (though well they may be!) but rather that they are composed of a large number of very small pieces, whose individual contributions to the whole we do not have the ability to investigate in a degree of detail sufficient to calculate the total cost precisely. It is much more efficient for us to recognize that virtually all components of cost are simply "uncertain" and to find some way to assign probabilities to various possible ranges of costs.*
>
> *An analogue is the situation in coin tossing where, in theory, if we knew all the physics involved and solved all the differential equations, we could predict with certainty whether a coin would fall "heads" or "tails". However, the combination of influences acting on the coin are too complicated to understand in sufficient detail to calculate the physical parameters of the coin's motion. So we do the next best thing: we bet that the uncertainties will probably average out in such a way that the coin will fall "heads" half the time and "tails" the other half. It is much more efficient to consider the deterministic physical process of coin tossing to be a "random" statistical process and to assign probabilities of 0.50 to each of the two possible outcomes, heads or tails.*"

Practice Point 18: Risk versus Uncertainty — There is an important distinction between the terms risk and uncertainty. *Risk* is the chance of loss or injury. In a situation that includes favorable and unfavorable events, risk is the probability an unfavorable event occurs. *Uncertainty* is the indefiniteness about the outcome of a situation. We analyze uncertainty *for the purpose of measuring risk*. In systems engineering, the analysis might focus on measuring the risk of: failing to achieve performance objectives, overrunning the budgeted cost, or delivering the system too late to meet user needs.

Practice Point 19: Subjective Probability Assessments — Probability theory is a well-established formalism for quantifying uncertainty. Its application to real-world systems cost analysis problems often involves the use of subjective probabilities. Subjective probabilities are those assigned to events on the basis of experience or judgment. They are measures of a person's degree-of-belief that an event will occur. Subjective probabilities are associated with one-time, non-repeatable events — those whose

[26] Book, S. A. 1997. Cost Risk Analysis – A Tutorial. *Risk Management Symposium Proceedings*. Los Angeles, California: The Aerospace Corporation.

probabilities cannot be objectively determined from a sample space of outcomes developed by repeated trials, or experimentation. Subjective probabilities must be consistent with the axioms of probability. For instance, if an engineer assigns a probability of 0.70 to the event "*the number of gates for the new processor chip* ***will not exceed*** *12000*," then it must follow the chip *will exceed* 12000 gates with probability 0.30. Subjective probabilities are *conditional* on the state of the person's knowledge, which changes with time.

To be credible, subjective probabilities should *only* be assigned to events by subject matter experts — persons with significant experience with events similar to the one under consideration. Instead of assigning a single subjective probability to an event, subject experts often find it easier to describe a function that depicts a distribution of probabilities. Such a distribution is sometimes called a *subjective probability distribution*. Subjective probability distributions are governed by the same mathematical properties of probability distributions associated with discrete or continuous random variables. Subjective probability distributions are most common in cost uncertainty analysis, particularly on the input-side of the process.

Practice Point 20: Correlation — Correlation is a necessary consideration in cost uncertainty analysis. It can exist between the costs of work breakdown structure (WBS) cost elements. Correlation can also exist between the cost of a cost element and the variables (e.g., weight, schedule) that define its cost.

Statistical theory offers a number of ways to measure correlation. Two popular measures are Pearson's product-moment correlation and Spearman's rank correlation. Subtleties concerning these measures must be understood to avoid errors in a cost uncertainty analysis. Pearson's product-moment correlation measures *linearity* between two random variables. Spearman's rank correlation measures their *monotonicity*. Thus, these two measures of correlation *can be very different*. Furthermore, the variance of a sum of random variables is a function of Pearson's product-moment correlation, *not* Spearman's rank correlation. Thus, from a WBS perspective, Pearson's product-moment correlation is the *only correct* measure of correlation to use when computing the variance of a sum of WBS element costs.

In cost uncertainty analysis, care must be taken if it is necessary to subjectively specify Pearson correlations. Pearson correlations can be restricted to a *subinterval* of -1 to +1 for random variables characterized by certain types of distribution functions [Garvey, 2000, Chapter 7]. Thus, the Pearson correlation between any two random variables cannot be assigned a value in a completely arbitrary way. If it is necessary to subjectively specify Pearson correlations, the reader should review the recently published work of Lurie-Golberg[27].

In practice, it is recommended that analysts express associations within the WBS through functional relationships (cost equations), as illustrated in this document. This allows the Pearson correlations *implied by these relationships* to be captured in the overall analysis. Pearson correlations that originate from logically defined functional relationships are more easily defended in cost reviews than those assigned by subjective assessments.

Practice Point 21: Capturing Cost-Schedule Uncertainties — Decision-makers require understanding how uncertainties between cost and schedule interact. A decision-maker might bet on a "high-risk" schedule in hopes of keeping cost within requirements. On the other hand, the decision-maker may be willing to assume "more cost" for a schedule with a small chance of being exceeded. This is a common tradeoff faced by decision-makers on systems engineering projects. The family of distributions in Garvey

[27] Lurie, P. M., and M. S. Goldberg. 1998. An Approximate Method for Sampling Correlated Random Variables from Partially-Specified Distributions. *Management Science*, Vol. 44, No. 2, pp. 203-218.

(2000) provides an analytical basis for computing this tradeoff, using joint and conditional cost-schedule probabilities. This family is a set of mathematical models that might be *hypothesized* for capturing the joint interactions between cost and schedule.

Practice Point 22: Approximating the Distribution Function of a System's Total Cost — Cost analysts are encouraged to *study the mathematical relationships* they define in a system's work breakdown structure, to see whether analytical approximations to the distribution function of $Cost_{WBS}$ can be argued. This enables the method of moments to be used as the primary risk analysis approach.

Analytical approximations can reveal much information about the "cost-behavior" in a system's WBS. The normal distribution often approximates the distribution function of $Cost_{WBS}$. There are many reasons for this approximation. Primary among them is that $Cost_{WBS}$ is a summation of WBS element costs. It is typical to have a mixture of independent and correlated WBS element costs within a system's WBS. Because of the Central Limit Theorem, the greater the number of independent WBS element costs the more it is that the distribution function of $Cost_{WBS}$ is approximately normal. The Central Limit Theorem is very powerful. It does not take many independent WBS element costs for the distribution function of $Cost_{WBS}$ to move towards normality. Such a move is evidenced when (1) a sufficient number of independent WBS element costs are summed and (2) when no element has a cost distribution with a much larger standard deviation than the standard deviations of the other element cost distributions. When conditions in the WBS result in $Cost_{WBS}$ being positively skewed (i.e., a non-normal distribution function), then the lognormal often approximates the distribution function of $Cost_{WBS}$.

Monte Carlo simulation is another approach for developing an empirical approximation to the distribution function of $Cost_{WBS}$. The Monte Carlo method is often needed when a system's WBS contains cost estimating relationships too complex for a strict method of moments approach. In Monte Carlo simulations, a question frequently asked is "*How many trials are necessary to have confidence in the output of the simulation*?" As a guideline, 10,000 trials (Monte Carlo samples) should be sufficient to meet the precision requirements for most Monte Carlo simulations in cost risk analysis.

Practice Point 23: Benefits of Cost Uncertainty Analysis — Cost uncertainty analysis provides decision-makers many benefits and important insights. These include:

Establish a Cost and Schedule Risk Baseline – Baseline probability distributions of program cost and schedule should be developed for a given system configuration (its technical baseline), acquisition strategy, and cost-schedule estimation approach. The baseline provides decision-makers visibility into potentially high-payoff areas for risk reduction initiatives. Baseline distributions assist in determining a program's cost and schedule that simultaneously have a specified probability of not being exceeded. They also provide decision-makers an assessment of the chance of achieving a budgeted (or proposed) cost and schedule, or cost for a given feasible schedule.

Measure Cost Risk – Cost risk analysis provides a basis for measuring the overall cost risk inherent to a program as a function of its specific uncertainties. This can be measured by difference between the program point estimate cost and the cost at a predefined confidence level, as set by budgetary decisions or management policies.

Conduct Risk Reduction Tradeoff Analyses – Cost risk analyses can be conducted to study the payoff of implementing risk reduction initiatives on lessening a program's cost, schedule, and performance risks. Families of probability distribution functions, as shown in Figure 4, can be generated to compare the cost and cost risk impacts of competing design options or acquisition strategies.

Document Program Risks and Risk Analysis Inputs – The validity and influence of any cost risk analysis relies on the engineering and cost team's experience, judgment, and knowledge of their program's risks and uncertainties. Documenting the team's insights into these considerations is a critical part of the process. Without it, the veracity of the cost risk analysis is easily questioned. Details about the analysis methodology, especially assumptions, are important to document. The methodology must be technically sound, traceable, and offer value-added problem structure and insights otherwise not visible. Decisions that successfully reduce or eliminate risk ultimately rest on human judgment. At best, this is aided by, not directed by, the methods in this paper.

4.1 HISTORICAL COST RISK DATA

Mentioned earlier, the Naval Center for Cost Analysis (NCCA) developed a large database of historical cost growth factors (CGF) from numerous major defense programs[28]. With this, the cost community now has evidence from program cost growth histories to guide the reasonableness of cost risk analyses.

The NCCA CGF database derives from Selected Acquisition Reports (SARs). The historical CGFs and CVs are from five different inputs: (1) commodity, (2) life cycle phase, (3) milestone, (4) inflation, and (5) quantity. The computed CVs are those of the cost growth factors, not the CVs of cost. The following tables present various summaries of the historical cost risk data derived from the NCCA study.

Cost Growth Factors & CVs for All DON MDAPs at MS B for 1969 & Later; n = 50				
	(Without Qty Adjustment)		(Quantity Adjusted)	
Statistics	Base-Year$	Then-Year$	Base-Year$	Then-Year$
Mean	1.48	1.84	1.23	1.36
Standard Deviation	0.94	1.60	0.44	0.69
CV	0.63	0.87	0.36	0.51

Cost Growth Factors & CVs for All DON MDAPs at MS B for 1980 & Later; n = 37				
	(Without Qty Adjustment)		(Quantity Adjusted)	
Statistics	Base-Year$	Then-Year$	Base-Year$	Then-Year$
Mean	1.34	1.37	1.22	1.23
Standard Deviation	0.87	0.94	0.46	0.49
CV	0.65	0.69	0.37	0.40

Cost Growth Factors & CVs for All DON MDAPs at MS B for 1990 & Later; n = 22				
	(Without Qty Adjustment)		(Quantity Adjusted)	
Statistics	Base-Year$	Then-Year$	Base-Year$	Then-Year$
Mean	1.46	1.51	1.36	1.36
Standard Deviation	0.74	0.81	0.41	0.45
CV	0.51	0.53	0.30	0.33

Table 8. NCCA Historical Cost Data Study: Cost Growth Factors & CVs

[28] The collection and analysis of historical data for use in cost risk analysis was under the auspices of Ms. Wendy Kunc, (Deputy Assistant Secretary for Cost and Economics, Office of the Assistant Secretary of the Navy (Financial Management & Comptroller) and Executive Director of the Naval Center for Cost Analysis (NCCA). The analysts who led statistical analysis of these data were Dr. Brian Flynn, Mr. Peter Braxton, and Mr. Richard Lee of Technomics, Inc. The reader is also directed to https://www.ncca.navy.mil/tools/tools.cfm for further information and access to the S-curve tool that contains the NCCA historical cost risk data.

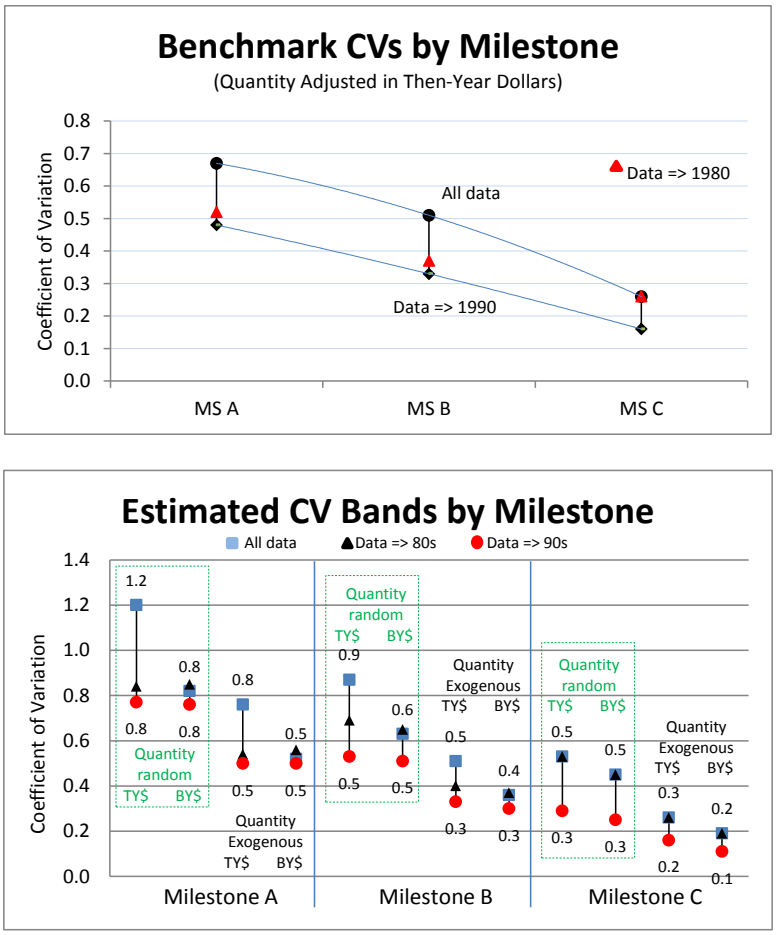


Figure 30. Historical CV Benchmark Bands[29]

Practice Point 24: Historical Data on Point Estimate Confidence

The historical cost growth data developed by the NCCA enabled deriving insights into point estimate confidence by major acquisition milestone. For example, an analysis of Milestone B program histories indicates a probability of 0.34 that program point estimates will not be exceeded. Discussed in Section 3.2.3 (Figure 29) this is the value α such that $P(Cost \le PE) = \alpha$. This is the first time an historical, data-driven, insight into point estimate confidence has been derived. It furthers the otherwise anecdotal experience that α often falls in the interval $0.10 \le \alpha \le 0.50$ for programs in these life cycle phases.

Although this finding is indicative of experience, more data and analyses along these lines should be collected and conducted on program cost histories across defense programs. This will provide the cost analysis community further empirical findings into cost growth factors, CVs across life cycle milestones, and point estimate probabilities. Deriving and documenting historical point estimate probabilities provides needed data-driven insights and advances cost realism in challenging acquisition environments.

[29] Sources: (1) NCCA Historical Cost Data Study, https://www.ncca.navy.mil/tools/tools.cfm (2) Garvey, P. R., Flynn, B. J., Braxton, P., Lee, R., 2012. "Enhanced Scenario-Based Method for Cost Risk Analysis: Theory, Application, and Implementation", *Journal of Cost Analysis and Parametrics*, Volume 5, Issue 2, pp. 98-142.

References and Related Literature

Arena, M. V., et al. 2006. *Impossible Certainty: Cost Risk Analysis for Air Force Systems*, Project Air Force, published by The RAND Corporation.

Book, Stephen A. 1995. "Fictions We Live By" The Aerospace Corporation, Society of Cost Estimating and Analysis.

Book, Stephen A. 1999. "Why Correlation Matters in Cost Estimating", The Aerospace Corporation, 32nd Annual DOD Cost Analysis Symposium (DODCAS).

Braxton, P. J., Flynn, B. J., Lee, R. C. 2012. "Enhanced Scenario-Based Method for Cost Risk Analysis: Theory, Application, and Implementation", 2012 SCEA/ISPA Joint Annual Conference and Training Workshop.

Dienemann, P. F. 1966. "Estimating Uncertainty Using Monte Carlo Techniques", RM-4854-PR. Santa Monica, California: The RAND Corporation.

Fisher, G. H. 1971. *Cost Considerations in Systems Analysis*. New York: American Elsevier Publishing Company, Inc.

GAO 2009. *Cost Estimating and Assessment Guide: Best Practices for Developing and Managing Capital Program Costs*, GAO-09-3SP, March 2009.

GAO 2013. *Assessments of Selected Weapon Programs*, GAO-13-294SP.

Garvey, P. R., 2000. *Probability Methods for Cost Uncertainty Analysis: A Systems Engineering Perspective*, Chapman-Hall/CRC Press, Taylor & Francis Group (UK), Boca Raton, London, New York, (2nd edition in-progress).

Garvey, P. R. 2008. "A Scenario-Based Method for Cost Risk Analysis", *Journal of Cost Analysis and Parametrics*, Vol. 1.

Garvey, P. R., Flynn, B. J., Braxton, P., Lee, R., 2012. "Enhanced Scenario-Based Method for Cost Risk Analysis: Theory, Application, and Implementation", *Journal of Cost Analysis and Parametrics*, Volume 5, Issue 2, pp. 98-142.

Goldberg, M. S., Lurie, P. M., 1998. "An Approximate Method for Sampling Correlated Random Variables From Partially-Specified Distributions", *Management Science*, Volume 44, Issue 2.

Hitch, C. J. 1955. *An Appreciation of Systems Analysis*, Santa Monica, CA: The RAND Corporation.

Missile Defense Agency (MDA), 2012. *Cost Estimating and Analysis Handbook*.

Morgan, M. G., M. Henrion. 1990. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. New York: Cambridge University Press.

NASA 2008. *Cost Estimating Handbook.*

Public Law 111-23, 2009. "Weapon Systems Acquisition Reform Act of 2009".

Rubinstein, R. Y. 1981. *Simulation and the Monte Carlo Method*. New York: John Wiley & Sons, Inc.

United States Air Force, Air Force Cost Analysis Agency, 2007. *Cost Risk and Uncertainty Analysis Handbook (CRUH).*
United States Air Force, Air Force Cost Analysis Agency, 2013. *Joint Cost-Schedule Risk and Uncertainty Handbook*.

USD(AT&L) Memorandum for Secretaries of the Military Departments and Directors of the Defense Agencies, 3 November 2010, "Implementation Directive for Better Buying Power: Obtaining Greater Efficiency and Productivity in Defense Spending".

Young, P. H. 1995. "A Generalized Probability Distribution for Cost/Schedule Uncertainty in Risk Assessment", *Proceedings of the 1995 Western Multi-Conference on Business/MIS Simulation Education*, The Society for Computer Simulation, San Diego, California.

Younossi, O., et al. 2007. *Is Weapon System Cost Growth Increasing?: A Quantitative Assessment of Completed and Ongoing Programs*, 2007, The RAND Corporation, ISBN 978-0-8330-4135-7.

APPENDIX A
POST-SIMULATION ADJUSTMENT TECHNIQUE (PSAT)

Appendix A illustrates the Post-Simulation Adjustment Technique (PSAT) through each implementation step. Applying the normal and lognormal distributions in the PSAT process is described.

Let $Cost_{WBS}$ be the sum of three WBS element cost random variables given by Equation A-1.

$$Cost_{WBS} = X_1 + X_2 + X_3 \tag{A-1}$$

Suppose WBS element costs X_1, X_2, and X_3 are determined by the relationships below.

$$X_1 \sim Unif(5,10) \tag{A-2}$$

$$X_2 = \frac{1}{2}X_1 + \frac{1}{8}X_3 \tag{A-3}$$

$$X_3 \sim Unif(20,30) \tag{A-4}$$

The uncertainties associated with X_1 and X_3 are represented by uniform distributions. WBS element cost X_2 is a function of X_1 and X_3. From a cost analysis perspective, X_2 is representative of a cost estimating heuristic – WBS element cost X_2 is equal to 50 percent of WBS element cost X_1 and 12.5 percent of WBS element cost X_3.

It follows that functional correlations exist between the pair (X_1, X_2) and the pair (X_2, X_3) but not between the pair (X_1, X_3). However, despite the absence of functional correlation between X_1 and X_3 suppose engineers and analysts know from experience their costs have some level of positive correlation. The following illustrates applying PSAT to capture these considerations in a Monte Carlo simulation, with Pearson product-moment correlations.

Step 1. Build and Implement a WBS Monte Carlo Simulation
Build and implement a Monte Carlo simulation within the WBS cost model using the same cost equations and estimating relationships that produced the overall point estimate of total program cost. The simulation will automatically capture all Pearson correlations between WBS element costs that stem from functional relationships between them, as they were defined to produce the original point estimate.

Figure A-1 shows a Monte Carlo simulation of $Cost_{WBS}$, given by Equation A-1, using the Crystal Ball tool. The table shown on the left is a snapshot of the raw results generated by the simulation. Shown are the first 45 trials of the simulation with its sample size set at 10,000 trials. In Figure A-1, Column A is the trial number. Column B is the value of $Cost_{WBS}$ for the indicated trial number, given the randomly sampled values for X_1, X_2, and X_3 shown by Columns C, D, and E, respectively. After 10,000 trials the Monte Carlo simulation of $Cost_{WBS}$ revealed that

$$E(Cost_{WBS})_{Sim} = 39.38 \tag{A-5}$$

$$Var(Cost_{WBS})_{Sim} = 15.2 \tag{A-6}$$

The simplicity of this example makes $E(Cost_{WBS})_{Sim}$ and $Var(Cost_{WBS})_{Sim}$ easy to mathematically derive by the method of moments. However, they can always be empirically determined by Monte Carlo simulation. Simulation is often the only way to approximate $E(Cost_{WBS})$ and $Var(Cost_{WBS})$ in cases with models characterized by complex cost equations and intricate work breakdown structures.

10,000 Trials (Monte Carlo Samples)

	A	B	C	D	E
	Trial values	**Sum**	**X2**	**X1**	**X3**
	1	45.77	8.11	9.06	28.59
	2	35.15	5.82	5.74	23.60
	3	37.96	6.60	7.15	24.21
	4	35.29	6.40	7.43	21.46
	5	44.69	8.02	9.15	27.53
	6	31.96	5.24	5.08	21.64
	7	33.66	5.89	6.46	21.31
	8	36.27	6.84	8.42	21.02
	9	33.22	5.50	5.42	22.30
	10	35.32	6.76	8.50	20.06
	11	44.92	8.31	9.96	26.65
	12	39.95	6.94	7.50	25.51
	13	36.64	7.17	9.30	20.17
	14	37.20	7.28	9.45	20.47
	15	36.21	6.03	6.03	24.16
	16	37.25	5.81	5.02	26.42
	17	31.55	5.30	5.38	20.87
	18	37.42	6.93	8.31	22.18
	19	36.43	6.12	6.23	24.08
	20	41.29	7.05	7.40	26.83
	21	47.66	8.58	9.86	29.22
	22	35.46	6.74	8.39	20.34
	23	35.34	6.01	6.26	23.07
	24	43.12	8.11	9.97	25.04
	25	40.71	7.32	8.39	25.01
	26	38.62	6.06	5.32	27.24
	27	43.92	7.25	7.10	29.58
	28	36.58	6.22	6.47	23.90
	29	36.02	5.79	5.37	24.86
	30	34.48	6.38	7.66	20.44
	31	32.98	5.63	5.88	21.47
	32	41.60	6.66	6.11	28.83
	33	39.23	6.82	7.40	25.01
	34	42.98	7.68	8.71	26.59
	35	39.60	7.19	8.38	24.03
	36	39.84	6.29	5.60	27.95
	37	44.09	7.44	7.61	29.05
	38	44.87	7.56	7.73	29.58
	39	41.91	7.80	9.42	24.69
	40	39.42	7.44	9.19	22.79
	41	43.06	7.06	6.83	29.16
	42	35.40	6.75	8.45	20.20
	43	34.64	5.52	5.02	24.09
	44	37.32	6.52	7.11	23.69
	45	41.28	7.51	8.76	25.01
	•	•	•	•	•
	•	•	•	•	•
10,000	•	•	•	•	•

Pearson Product-Moment Correlations

(X2, X1)	(X2, X3)
0.895	0.439

= PEARSON(C1:C10001, E1:E10001)

= PEARSON(C1:C10001, D1:D10001)

Cost Element X1 = 7 ($M)

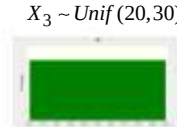
$X_1 \sim Unif\,(5,10)$

Cost Element X2 = 6.625 ($M)

$$X_2 = \frac{1}{2}X_1 + \frac{1}{8}X_3$$

Cost Element X3 = 25 ($M)

$X_3 \sim Unif\,(20,30)$



$$Sum = X_1 + X_2 + X_3$$

Figure A-1. A Monte Carlo Simulation of $Cost_{WBS} = X_1 + X_2 + X_3$

The mean and variance statistics below can be derived (in this example) by the method of moments or extracted from the outputs generated by the Monte Carlo simulation. The two correlation statistics shown below were computed in Excel from the extracted simulation results as illustrated in Figure A-1.

$$E(X_1) = 7.5,\ E(X_2) = 6.875,\ E(X_3) = 25$$

$$Var(X_1) = 25/12,\ Var(X_2) = 0.65104,\ Var(X_3) = 100/12$$

$$\rho_{X_1,X_2} = 0.895, \text{ and } \rho_{X_2,X_3} = 0.439$$

Given the above, the values for $E(Cost_{WBS})$ and $Var(Cost_{WBS})$ can also be computed by the method of moments and compared to the simulation values; that is, given $Cost_{WBS} = X_1 + X_2 + X_3$ then

$$E(Cost_{WBS}) = E(X_1 + X_2 + X_3) = E(X_1) + E(X_2) + E(X_3) \tag{A-7}$$

$$Var(Cost_{WBS}) = Var(X_1 + X_2 + X_3) = Var(X_1) + Var(X_2) + Var(X_3) + 2\rho_{X_1,X_2}\sigma_{X_1}\sigma_{X_2} + 2\rho_{X_2,X_3}\sigma_{X_2}\sigma_{X_3} \tag{A-8}$$

$$E(Cost_{WBS}) = E(X_1) + E(X_2) + E(X_3) = 7.5 + 6.875 + 25 = 39.38 \approx E(Cost_{WBS})_{Sim} \tag{A-9}$$

$$\begin{aligned} Var(Cost_{WBS}) &= Var(X_1) + Var(X_2) + Var(X_3) + 2\rho_{X_1,X_2}\sigma_{X_1}\sigma_{X_2} + 2\rho_{X_2,X_3}\sigma_{X_2}\sigma_{X_3} \\ &= (25/12) + 0.65104 + (100/12) \\ &\quad + 2(0.895)\sqrt{25/12}\sqrt{0.65104} + 2(0.439)\sqrt{0.65104}\sqrt{100/12} \\ &= 15.2 \approx Var(Cost_{WBS})_{Sim} \end{aligned} \tag{A-10}$$

Allowing for numerical rounding, and minor sampling errors always present in simulations, the method of moments results given by Equations A-9 and A-10 are consistent with the empirical findings from the Monte Carlo simulation given by Equations A-5 and A-6.

Step 2. Make a Post-Simulation Adjustment to the Simulated Variance
In this example, WBS element costs X_1 and X_3 were determined in a way that X_1 was not computed by a cost estimating relationship defined by a function of X_3 (and vice-versa). Thus, the pair of WBS element costs (X_1, X_3) has no functional correlation between them. The Monte Carlo simulation will treat X_1 and X_3 as uncorrelated. This is equivalent to setting $\rho_{X_1,X_3} = 0$ in the method of moments.

However, suppose engineers and analysts know from experience the costs of cost elements X_1 and X_3 have some level of positive correlation. This correlation must also be captured in the cost risk analysis. Since correlation only affects $Var(Cost_{WBS})$, a post-simulation adjustment to Equation A-10 is needed. To do this, PSAT defines an "adjusted" variance $adjVar(Cost_{WBS})$ as follows:

$$adjVar(Cost_{WBS}) = Var(Cost_{WBS})_{Sim} + \overbrace{\underbrace{2\rho_{X_1,X_3}\sigma_{X_1}\sigma_{X_3}}_{\text{Additional covariance needed to capture the Pearson product-moment correlation between } X_1 \text{ and } X_3}}^{Covariance(X_1, X_3) =} \tag{A-11}$$

In Equation A-11, the term $Var(Cost_{WBS})_{Sim}$ is produced by the Monte Carlo simulation (Step 1, Equation A-6). The variables σ_{X_1} and σ_{X_3} are the square roots of the $Var(X_1)$ and $Var(X_3)$, respectively. Their values are also produced by the Monte Carlo simulation (Step 1). In Equation A-11, the variable ρ_{X_1,X_3} is the Pearson product-moment correlation of the pair (X_1, X_3). The value for ρ_{X_1,X_3} can originate from historical data, subject experts, or assigned from guidelines in Table 3.

In this example, suppose the engineers and analysts felt $\rho_{X_1,X_3} = 1/4$ is a good reflection of the strength and direction of this correlation between X_1 and X_3. Using Equation A-11, it follows that:

$$adjVar(Cost_{WBS}) = Var(Cost_{WBS})_{Sim} + 2(1/4)\sqrt{25/12}\sqrt{100/12} \tag{A-12}$$

$$adjVar(Cost_{WBS}) = 15.2 + 2(1/4)\sqrt{25/12}\sqrt{100/12} = 17.3 \tag{A-13}$$

Equation A-13 is the post-simulation adjustment to $Var(Cost_{WBS})_{Sim}$ for to capture $\rho_{X_1,X_3} = 1/4$.

Step 3. Make a Post-Simulation Adjustment to the Simulated Probability Distribution of $Cost_{WBS}$
With Step 2 completed, the simulated probability distribution of $Cost_{WBS}$ must then be adjusted. This is necessary to account for the adjusted variance $adjVar(Cost_{WBS})$ determined in Step 2.

To adjust the simulated probability distribution of $Cost_{WBS}$, one can reasonably assume the modified distribution is either normal or lognormal with mean $E(Cost_{WBS})_{Sim}$ and variance $adjVar(Cost_{WBS})$. Figure A-2 presents the adjusted distribution of $Cost_{WBS}$ assuming it is represented by a normal distribution with mean 39.38 and variance 17.3. In Figure A-2, the solid red line is the assumed normal. The solid blue line is the adjusted distribution of $Cost_{WBS}$ assuming it is represented by a lognormal distribution with mean 39.38 and variance 17.3.

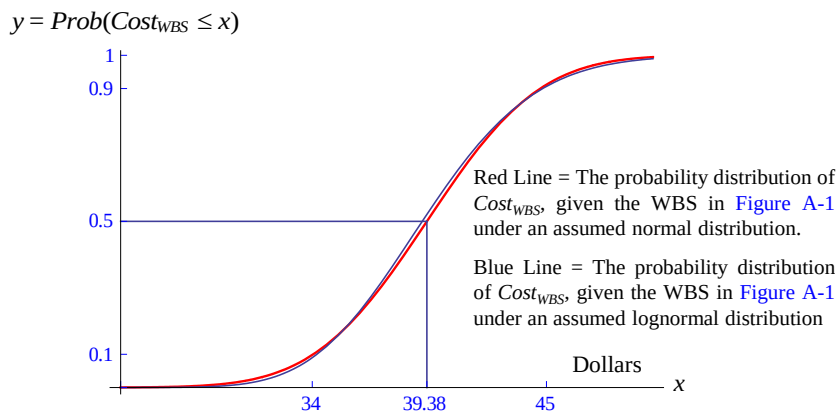


Figure A-2. Post-Simulation Adjusted Probability Distributions for $Cost_{WBS}$

APPENDIX B
METHOD OF MOMENTS WBS EXAMPLE

Suppose the cost of an electronic system is represented by the work breakdown structure in Table B-1. This WBS consists of element costs $X_1, X_2, X_3, \ldots, X_{10}$. Suppose

$$Cost_{WBS} = X_1 + X_2 + X_3 + \ldots + X_{10} \tag{B-1}$$

and the random variables $X_1, W, X_5, X_7, X_8, X_9$ in Table B-1 are independent – they are uncorrelated. Use the method of moments to determine the following:

A) $E(Cost_{WBS})$ and $Var(Cost_{WBS})$.

B) A probability distribution function that approximates the distribution of $Cost_{WBS}$.

C) The value of $Cost_{WBS}$ that has a 95 percent chance of not being exceeded.

Cost Element Name	Cost Element Cost X_i ($M)	Distribution of X_i or the Applicable Functional Relationship
Prime Mission Product (PMP)	X_1	$N(12.5, 6.6)$
System Engineering and Program Management (SEPM)	X_2	$X_2 = \frac{1}{2} X_1$
System Test & Evaluation (STE)	X_3	$X_3 = \frac{1}{4} X_1 + \frac{1}{8} X_2 + W$, where $W \sim Unif(0.6, 1.0)$
Data and Technical Orders	X_4	$X_4 = \frac{1}{10} X_1$
Site Survey and Activation	X_5	$Trng(5.1, 6.6, 12.1)$
Initial Spares	X_6	$X_6 = \frac{1}{10} X_1$
System Warranty	X_7	$Unif(0.9, 1.3)$
Early Prototype Phase	X_8	$Trng(1.0, 1.5, 2.4)$
Operations Support	X_9	$Trng(0.9, 1.2, 1.6)$
System Training	X_{10}	$X_{10} = \frac{1}{4} X_1$

Table B-1. An Electronic System Work Breakdown Structure

Solution

A) Given Equation B-1 and the relationships in Table B-1, then Equation B-1 can be written as

$$Cost_{WBS} = X_1 + \frac{1}{2} X_1 + \left(\frac{1}{4} X_1 + \frac{1}{8} X_2 + W\right) + \frac{1}{10} X_1 + X_5 + \frac{1}{10} X_1 + X_7 + X_8 + X_9 + \frac{1}{4} X_1 \tag{B-2}$$

Combining and simplifying this expression yields the following:

$$Cost_{WBS} = \frac{181}{80} X_1 + W + X_5 + X_7 + X_8 + X_9 \tag{B-3}$$

$$E(Cost_{WBS}) = \frac{181}{80} E(X_1) + E(W) + E(X_5) + E(X_7) + E(X_8) + E(X_9) \tag{B-4}$$

$$Var(Cost_{WBS}) = \left(\frac{181}{80}\right)^2 Var(X_1) + Var(W) + Var(X_5) + Var(X_7) + Var(X_8) + Var(X_9) \tag{B-5}$$

The variance of $Cost_{WBS}$ given by Equation B-5 captures all Pearson product-moment correlations present in the WBS from the *cost estimating functional relationships* defined in Table B-1. Equation B-5 also reflects that $X_1, W, X_5, X_7, X_8,$, and X_9 were given to be uncorrelated.

To compute the mean and variance of $Cost_{WBS}$ given by Equations B-4 and B-5, respectively, the means and variances of $X_1, W, X_5, X_7, X_8,$, and X_9 (the terms in Equations B-4 and B-5) are needed. Table B-2 presents these statistics. They are determined by standard formulas available in Garvey (2000).

Cost Element Cost X_i (\$M)	$E(X_i)$ (\$M)	$Var(X_i)$ (\$M)^2
X_1	12.500	6.6
W	0.800	0.16/12
X_5	7.933	40.75/18
X_7	1.100	0.16/12
X_8	1.633	1.51/18
X_9	1.233	0.37/18

Table B-2. Cost Statistics for $X_1, W, X_5, X_7, X_8,$, and X_9

Substituting the values in Table B-2 into Equations B-4 and B-5 produces the mean and variance of the electronic system's total WBS cost, given by Equations B-6 and B-7, respectively.

$$E(Cost_{WBS}) = 40.98 \text{ (\$M)} \tag{B-6}$$

$$Var(Cost_{WBS}) = 36.18 \text{ (\$M)}^2 \tag{B-7}$$

$$\sigma_{Cost_{WBS}} = \sqrt{Var(Cost_{WBS})} = 6.015 \text{ (\$M)} \tag{B-8}$$

B) To approximate the distribution function of $Cost_{WBS}$, observe the following. First, the random variables $X_1, W, X_5, X_7, X_8,$, and X_9 are independent. Hence, the Central Limit Theorem will draw the shape of the distribution of $Cost_{WBS}$ towards a normal distribution. Second, the random variables $X_2, X_3, X_4, X_6,$ and X_{10} are highly correlated to X_1, which is normally distributed. With this, it can be shown that

$$\rho_{X_v, X_1} = 1 \quad (v = 2,4,6,10) \text{ and } \rho_{X_3, X_1} = 0.9898$$

Thus, it is reasonable to conclude (for this example) the probability distribution function of $Cost_{WBS}$ is approximately normal — with mean and variance given by Equations B-6 and B-7, respectively. Figure B-1 presents a graph of the probability distribution function of $Cost_{WBS}$.

C) From Figure B-1, it can be seen that $Cost_{WBS} = 50.87$ (\$M) has a 95 percent chance of not being exceeded. Thus, a cost equal to 50.87 (\$M) has only a 5 percent chance of being exceeded. Equivalently, 50.87 (\$M) is the 0.95-fractile of $Cost_{WBS}$; that is, $x_{0.95} = 50.87$ (\$M). If the electronic system is considering budgeting at the 95th percentile, then a cost reserve of nearly 10 (\$M) is needed above its cost at the 50th percentile for this high level of estimate confidence.

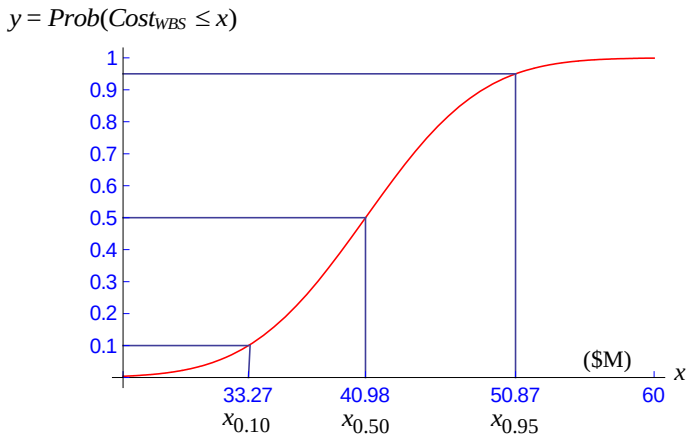


Figure B-1. Normal Probability Distribution for $Cost_{WBS}$

APPENDIX C
eSBM LOGNORMAL EXAMPLE

In the eSBM discussion, Example 1 illustrated its application when the distribution function of a program's total cost is normal. This appendix shows the same example, but illustrates the application of eSBM when the distribution function of a program's total cost is lognormal.

Example 1 (Lognormal Distribution)
Suppose the distribution function of a program's total cost is lognormal. Suppose the program's point estimate cost is $100M and this was assessed to fall at the 25th percentile. Suppose the type and life cycle phase of the program is such that 30 percent variability in cost around the mean has been historically seen. Suppose the protect scenario was defined and determined to cost $145M.

A) Compute the mean and standard deviation of $Cost_{WBS}$.
B) Determine the confidence level of the protect scenario cost and its associated cost reserve.

Solution
A) From Equations 32 and 33

$$a = \ln PE - z\sqrt{\ln(1+(CV)^2)} = \ln(100) - (-0.6745)\sqrt{\ln(1+(0.30)^2)} = 4.80317$$

$$b = \sqrt{\ln(1+(CV)^2)} = \sqrt{\ln(1+(0.30)^2)} = 0.29356$$

From Equations 30 and 31 the above mean and standard deviation are translated into dollar units.

$$\mu = e^{a+\frac{1}{2}b^2} = e^{4.80317+\frac{1}{2}(0.29356)^2} \approx 127.3 \text{ (\$M)}$$

$$\sigma = \sqrt{e^{2a+b^2}(e^{b^2}-1)} = \mu\sqrt{(e^{b^2}-1)}$$

$$= 127.3\sqrt{(e^{(0.29356)^2}-1)} \approx 38.2 \text{ (\$M)}$$

b) To determine the confidence level of the protect scenario we need to find α_{PS} such that

$$P(Cost \le PS = 145) = \alpha_{PS}$$

Finding α_{PS} is equivalent to solving $a + z_{PS}(b) = \ln PS$ for z_{PS} . From this,

$$z_{PS} = \frac{\ln PS - a}{b}$$

Since $PS = 145$, $a = 4.80317$, and $b = 0.29356$ it follows that

$$z_{PS} = \frac{\ln PS - a}{b} = \frac{\ln 145 - 4.80317}{0.29356} = 0.59123$$

Thus, we want α such that $P(Z \le z_{PS} = 0.59123) = \alpha$. Values for α are available in Excel as follows. With $z_{PS} = 0.59123$, enter into Excel: NORM.S.DIST(0.59123,TRUE); that is,

$$\alpha = \text{NORM.S.DIST}\left(z_{PS}, \text{TRUE}\right) = \text{NORM.S.DIST}\left(0.59123, \text{TRUE}\right) = 0.723$$

$$\Rightarrow P(Z \le z_{PS} = 0.59123) \approx 0.723$$

Therefore, the $145M protect scenario cost falls at the 72nd percentile of the distribution. This implies a cost reserve CR equal to $45M.

APPENDIX D[30]
UNRAVELING THE S-CURVE

One of the least documented topics in cost risk analysis is conveying its findings to decision-makers. The cost community's focus has been on the front end of the process (methods and practices) rather than on the back end, specifically, what the analysis reveals and how best to convey its findings. This section discusses one aspect of this topic by addressing the question "*I've generated an S-curve – what does it reveal about my program's cost risk and how should I present these findings to decision-makers*?"

What is the S-Curve?

The S-curve is an informal term for the probability distribution of the cost of a program. Figure D-1, illustrates two ways to present a probability distribution. One way is the probability density function (PDF), as shown on the left of Figure D-1. The other way is the cumulative distribution function (CDF), as shown on the right of Figure D-1. The CDF is informally called the S-curve.

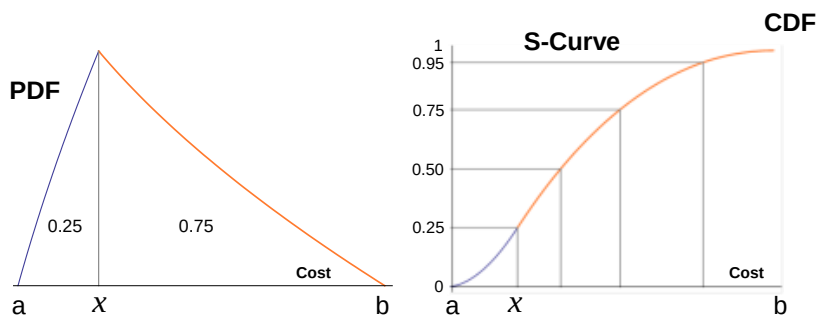


Figure D-1. Ways to View a Program's Cost Probability Distribution

In Figure D-1, the range of possible cost outcomes for a program is given by the interval $a \le x \le b$. These distributions reveal the confidence that the actual cost of a program will not exceed any cost in the range of possible outcomes. For example, the probability that the actual cost of the program will be less than or equal to *x* is 25 percent. In a PDF, this probability is the area under the curve. In a CDF, this probability is the value 0.25 along the vertical axis.

The PDF is the most common form of a probability distribution used to characterize the cost uncertainties of elements that comprise a program's work breakdown structure (WBS). This is shown in Figure D-2 by the elements shown on the left, which is the input side of a cost risk analysis. The right side of Figure D-2 Shows the outputs of a cost risk analysis, where the CDF or S-curve is the most common form used to express percentile levels of confidence that the actual cost of a program is less than or equal to a value *x*.

[30] The numerical example in this appendix was developed by Mr. Raymond P. Covert (rpcovert@covarus.com) President of Covarus, LLC (http://www.covarus.com/).

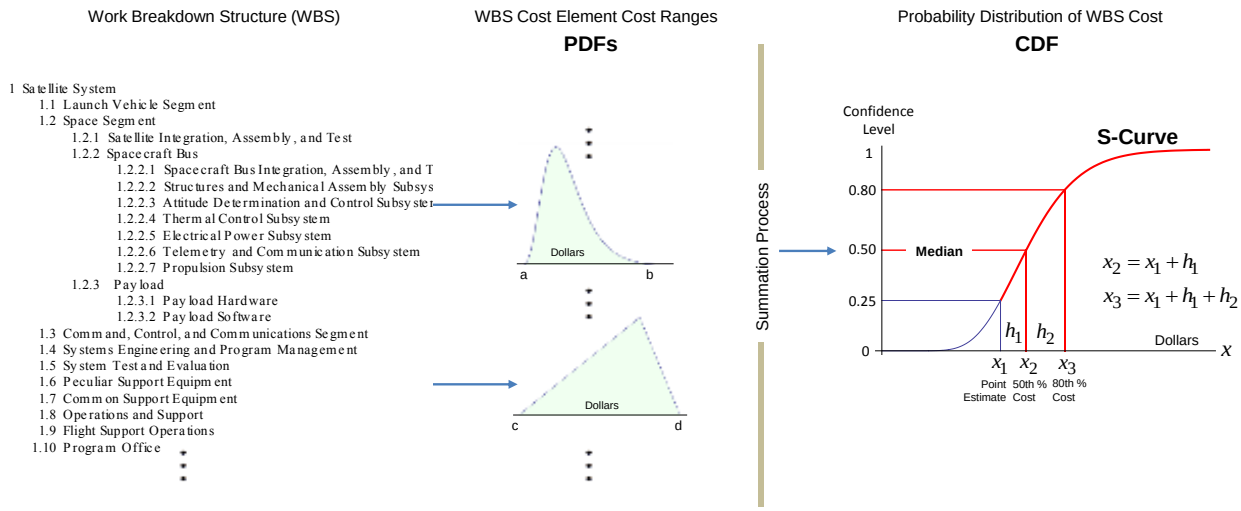


Figure D-2. Probability Distributions in WBS Cost Risk Analysis

Unraveling the S-Curve

The S-curve provides decision-makers a basis for examining tradeoffs between a program's point estimate cost[31] and its confidence level. For example, in Figure D-2, there is a 25 percent chance the actual program cost will be less than or equal to x_1 dollars, a 50 percent chance the actual program cost will be less than or equal to x_2 dollars, and an 80 percent chance the actual program cost will be less than or equal to x_3 dollars.

The variance of the S-curve, in Figure D-2, affects the amount of risk dollars needed to budget a program at a given confidence level, relative to (say) the program's point estimate cost. For example, in Figure D-2, if a program is budgeted to the 50th percentile cost x_2, then relative to x_1 there are h_1 risk dollars contained in x_2. If a program is budgeted to the 80th percentile cost x_3, then relative to x_1 there are $(h_1 + h_2)$ risk dollars contained in x_3. Thus, the risk dollars between x_1 and cost outcomes greater than x_1 is from the accumulation of cost risk from individual WBS elements that comprise a program's total cost. How can the S-curve be unraveled to reveal which WBS elements drive the amount of risk dollars needed for a given confidence? The following algorithm is used to address this question.

Book's Algorithm[32]

Book's algorithm is designed to allocate the risk dollars of a program into the individual WBS elements as a function of the cost risk of each element. The total risk dollars of a program is the difference between its point estimate cost and its cost at a confidence level greater than that of the point estimate, such as the dollars given by h_1 or $(h_1 + h_2)$ in Figure D-2. WBS elements allocated the largest fraction of risk dollars are the cost risk driving elements of the program and signal potential priorities for risk management actions. The following are the key terms and equations of Book's algorithm. A numerical example of the algorithm is then presented.

[31] For purposes of this paper, the point estimate (PE) cost is the cost that does not include allowances for cost uncertainty. The PE cost is the sum of the WBS element costs summed across a program's work breakdown structure without adjustments for uncertainty. The PE cost is often developed from a program's cost analysis requirements description document (CARD).

[32] This algorithm was created by S. A. Book, Ph.D. It is published in the 2008 NASA Cost Estimating Handbook, NASA Cost Analysis Division, http://www.nasa.gov/offices/ipce/CA.html.

From Figure D-2, suppose a program is budgeted to the 80th percentile cost x_3. Then, relative to x_1 there are $(h_1 + h_2)$ risk dollars contained in x_3. To allocate $(h_1 + h_2)$ to the individual WBS elements, define the *need* of element *k* as the difference between its 80th percentile cost and its point estimate cost; that is,

$$Need_k = x_{k,0.80} - PE_k \tag{1}$$

where $x_{k,0.80}$ is the 80th percentile cost of WBS element *k* and PE_k is its point estimate cost. Equation D-1 is set equal to zero if $x_{k,0.80} \le PE_k$ plus any correlation effects due the impacts of the needs of other WBS elements with which element *k* is correlated. Equation D-1 is the above average portion of σ_k measuring only the possible shortfall in dollars for WBS element *k*, if any of the identified risk associated with this element are realized. The fraction of risk dollars to be allocated to WBS element *k*, including correlation effects, is given by

$$Alloc_k = \frac{\text{Correlated Need of WBS Element } k}{\text{Total Need Base}} = \frac{\left(\sum_{j=1}^{n} \rho_{jk} Need_j Need_k\right)}{\sum_{k=1}^{n} \sum_{j=1}^{n} \rho_{jk} Need_j Need_k} \tag{2}$$

Numerical Example

Consider the WBS on the left in Table D-1. Suppose the cost probability distribution of each WBS element is lognormal with mean and standard deviation (sigma) given in Table D-1. Suppose a Monte Carlo simulation was run on the WBS that generated the first 20 of 5000 trials on the right in Table D-1.

WBS Element	Point Estimate ($M)	Mean ($M)	Sigma ($M)
1. Satellite	200	230	69
2. Launch	80	90	9
3. Ground	100	70	14
4. Data Distribution	300	350	140
Program Cost	680	740	

Monte Carlo Simulation Results (First 20 of 5000 Trials)					
Trial Number	1. Satellite	2. Launch	3. Ground	4. Data Distribution	Total ($M)
1	$ 205.07	$ 84.49	$ 59.97	$ 432.92	$ 782.44
2	$ 227.80	$ 86.78	$ 69.61	$ 482.22	$ 866.40
3	$ 159.15	$ 94.66	$ 64.89	$ 497.95	$ 816.66
4	$ 261.52	$ 76.97	$ 60.33	$ 151.53	$ 550.36
5	$ 302.11	$ 99.12	$ 56.54	$ 242.11	$ 699.88
6	$ 225.99	$ 101.40	$ 68.41	$ 368.03	$ 763.83
7	$ 335.86	$ 99.47	$ 82.75	$ 845.43	$ 1,363.51
8	$ 279.25	$ 98.92	$ 70.53	$ 443.87	$ 892.57
9	$ 160.00	$ 91.61	$ 82.53	$ 392.05	$ 726.18
10	$ 250.66	$ 87.61	$ 74.07	$ 280.87	$ 693.21
11	$ 197.80	$ 82.91	$ 60.02	$ 123.35	$ 464.08
12	$ 203.37	$ 90.22	$ 68.88	$ 360.09	$ 722.55
13	$ 400.09	$ 91.35	$ 68.83	$ 560.93	$ 1,121.21
14	$ 307.00	$ 100.34	$ 77.11	$ 254.32	$ 738.77
15	$ 310.87	$ 88.79	$ 67.98	$ 328.28	$ 795.91
16	$ 260.09	$ 91.97	$ 47.55	$ 224.12	$ 623.74
17	$ 149.99	$ 72.46	$ 82.96	$ 306.97	$ 612.38
18	$ 251.13	$ 78.63	$ 58.47	$ 215.72	$ 603.95
19	$ 259.10	$ 100.79	$ 96.57	$ 491.69	$ 948.16
20	$ 259.51	$ 90.62	$ 91.14	$ 396.04	$ 837.30

Table D-1. A Program WBS and Monte Carlo Simulation (Partial Results)

The next step in the algorithm is to compute Pearson correlation coefficients between each pair of WBS element costs. This can be performed by (1) exporting the WBS element costs from all trials of the Monte Carlo simulation (the first 20 of 5000 trials are shown in Table D-1) then (2) computing Pearson

correlation coefficients for each column pair of WBS element costs from the exported values. The Excel function CORREL(*array1*, *array2*) can be used to perform this computation, where the columns of the WBS element costs shown in Table D-1 are the entries for *array1* and *array2*. The results of this operation are provided in Table D-2[33].

Correlation Matrix	1. Satellite	2. Launch	3. Ground	4. Data Distribution
1. Satellite	1	0.227186	0.252635523	0.320887907
2. Launch	0.227186	1	0.203265326	0.190462092
3. Ground	0.2526355	0.20326533	1	0.29252485
4. Data Distribution	0.3208879	0.19046209	0.29252485	1

Table D-2. Pearson Correlation Coefficients Computed From the Simulation Data in Table D-1

Suppose the program represented by the work breakdown structure in Table D-1, is required to be budgeted at the 80th percentile confidence level. Given this, Table D-3 presents the costs of the program and its WBS elements at this confidence level. The Excel function PERCENTILE.INC(*array*,0.80) can be used to compute these costs. For example, if *array* is set equal to the satellite column of cost data from the Monte Carlo simulation in Table D-1, then this Excel function produces 283.56 ($M) as the 80th percentile cost for the satellite WBS element. Equation D-1 is then applied to compute $Need_k$ as shown in Table D-3.

Costs ($M)	Point Estimate	80th Percentile	Need_k
1. Satellite	200	283.56	83.56
2. Launch	80	97.15	17.15
3. Ground	100	81.21	0.00
4. Data Distribution	300	447.84	147.84
Program Cost	680	878.48	198.48

Table D-3. Costs at 80th Percentile

From Table D-3 the fraction of risk dollars to be allocated to WBS element *k*, including correlation effects can now be computed from Equation D-2. For example, for the satellite WBS element (k = 1) we have

$$Alloc_1 = \frac{\text{Correlated Need of WBS Element 1}}{\text{Total Need Base}} = \frac{\left(\sum_{j=1}^{4} \rho_{j1} Need_j Need_1\right)}{\sum_{k=1}^{4}\sum_{j=1}^{4} \rho_{jk} Need_j Need_k}$$

where

$$\sum_{j=1}^{4} \rho_{j1} Need_j Need_1 = \rho_{11} Need_1 Need_1 + \rho_{21} Need_2 Need_1 + \rho_{31} Need_3 Need_1 + \rho_{41} Need_4 Need_1$$

$$= (1)(83.56)(83.56) + (0.227186)(17.15)(83.56) + (0.2526355)(0)(83.56) + (0.3208879)(147.84)(83.56) = 11272$$

and

$$\sum_{k=1}^{4}\sum_{j=1}^{4} \rho_{jk} Need_j Need_k = [83.56, 17.15, 0, 147.48] \begin{pmatrix} 1 & 0.227186 & 0.2526355 & 0.3208879 \\ 0.227186 & 1 & 0.20326533 & 0.19046209 \\ 0.2526355 & 0.20326533 & 1 & 0.29252485 \\ 0.3208879 & 0.19046209 & 0.29252485 & 1 \end{pmatrix} \begin{bmatrix} 83.56 \\ 17.15 \\ 0 \\ 147.48 \end{bmatrix} = 38678$$

[33] The number of significant digits in Table D-2 is for traceability in this example; in practice, 3 or 4 significant digits is best.

From this, it follows that

$$Alloc_1 = \frac{\left(\sum_{j=1}^{4} \rho_{j1} Need_j Need_1\right)}{\sum_{k=1}^{4} \sum_{j=1}^{4} \rho_{jk} Need_j Need_k} = \frac{11272}{38678} = 0.29$$

Therefore, the satellite WBS element requires 29 percent of the 198.48 ($M) risk dollars – given the cost of the program is to be budgeted at the 80th percentile confidence level. Similar computations made for the other WBS elements reveal the results shown in Figure D-3. From this, the following additional findings can be conveyed to the decision-maker:

- The Ground WBS element may be over-estimated since it needs no additional risk dollars.
- The largest cost risk driver of the program is the Data Distribution WBS element. This element consumes 68 percent of the risk dollars needed for the program to be budgeted at the 80th percentile.

The Data Distribution WBS element is a prime candidate for further management attention to reduce its high cost risk to the program. Its high demand for risk dollars is a consequence of its large cost uncertainty (seen by its sigma value in Table D-1) and its low point estimate cost. The risk dollars might be reduced simply by providing a more complete and less uncertain system definition combined with better cost estimating methods.

WBS Element	% Allocation of Risk Dollars	Allocation of Risk Dollars ($M)
1. Satellite	29.14	57.85
2. Launch	2.85	5.66
3. Ground	0	0.00
4. Data Distribution	68.01	134.98
Total	100	198.48

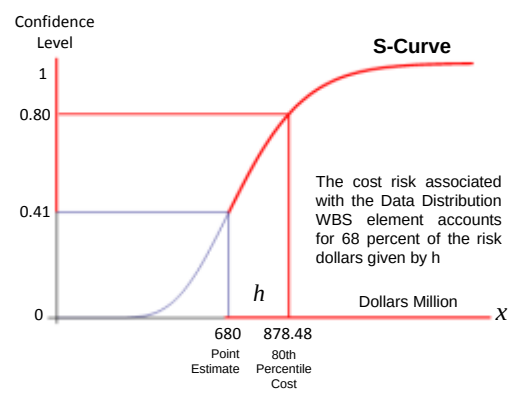


Figure D-3. Unraveling the S-Curve: Allocation of 80th Percentile Risk Dollars

Summary

This section discussed a key aspect of communicating findings from a cost risk analysis to decision-makers; that is, unraveling a program's S-curve to identify the elements of cost that drive the greatest amount of cost risk. In the preceding numerical example, it was shown that one WBS element more than others is driving the program's cost risk. Identifying cost risk drivers, as in the way presented, fosters risk reducing management actions to be taken as early as possible – such that program cost, schedule, and technical goals are achieved.