

An Acquisition Process for the Management of Risks of Cost Overrun and Time Delay Associated with Software Development

Y. Haimes
C. P. Chittister

Technical Report
CMU/SEI-93-TR-028

December 1993

TECHNICAL REPORT
CMU/SEI-93-TR-028

Software Risk Management Program
Unlimited distribution subject to the copyright.

This technical report was prepared for the

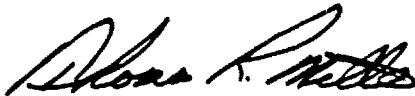
SEI Joint Program Office
ESC/ENS
Hanscom AFB, MA 01731-2116

The ideas and findings in this report should not be construed as an official DoD position. It is published in the interest of scientific and technical information exchange.

Review and Approval

This report has been reviewed and is approved for publication.

FOR THE COMMANDER



Thomas R. Miller, Lt Col, USAF
SEI Joint Program Office

The Software Engineering Institute is sponsored by the U.S. Department of Defense.
This report was funded by the U.S. Department of Defense.

Copyright © 1993 by Carnegie Mellon University.

This document is available through the Defense Technical Information Center. DTIC provides access to and transfer of scientific and technical information for DoD personnel, DoD contractors and potential contractors, and other U.S. Government agency personnel and their contractors. To obtain a copy, please contact DTIC directly. Defense Technical Information Center, Attn: FDRA, Cameron Station, Alexandria, VA 22304-6145

Copies of this document are also available through the National Technical Information Service. For information on ordering, please contact NTIS directly. National Technical Information Service, U.S. Department of Commerce, Springfield, VA 22161

Copies of this document are also available from Research Access, Inc., 800 Vinal Street, Pittsburgh, PA 15212, Telephone (412) 321-2992 or 1-800-685-6510, Fax (412) 321-2994

Use of any trademarks in this report is not intended in any way to infringe on the rights of the trademark holder.

Table of Contents

Preface	vii
Acknowledgements	ix
1. Introduction	3
2. Overview of the Conceptual Framework	5
2.1. Phase I	8
2.2. Phase II	8
2.2.1. Extreme Events	8
2.2.2. Contractor's Attitude	9
2.3. Phase III	9
2.4. Phase IV	10
3. Critical Factors That Affect Software Nontechnical Risk	11
3.1. The Customer	11
3.2. The Contractor(s)	12
3.3. The Interface Between the Customer and the Contractor(s)	12
3.4. The State of Technology and Knowhow	12
3.5. The Complexity of the Specification Requirements	13
3.6. The Add-On Modifications and Refinements	13
3.7. The Availability of Appropriate Resources	13
3.8. The Models Used for Project Cost Estimation and Schedule Projection	14
4. Basis for Variances in Cost Estimation	17
5. The Quantification of Software Nontechnical Risk and the Evaluation of Variances	19
6. Risk of Extreme Events	23
7. Problem 1	27
8. General Notation	35
9. Problem 2	37
10. Problem 3	45
11. Comparative Analysis Among Contractors	51
12. Epilogue	55
References	57

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability	
Dist	Avail and/or Special

List of Tables

Table 7-1 - Tabular CDF for Problem 1	28
Table 8-1 - Summary of Available Information	36
Table 9-1 - Comparative Tabular CDF (Problem 2)	41
Table 9-2 - Summary of Results (Problem 2)	42
Table 10-1 - Comparative Assessed Values for the Triangular Distribution (Problem 3)	47
Table 10-2 - Summary of Results for Triangular Distribution (Problem 3)	49

List of Figures

Figure 2-1 - Proposed Acquisition Process	7
Figure 6-1 - Mapping of the Probability Partitioning onto the Damage Axis	25
Figure 7-1 - Graphical CDF for Project Cost Increase (Problem 1)	29
Figure 7-2 - Probability Density Function for Project Cost Increase (Problem 1)	29
Figure 7-3 - Exceedance Probability for Project Cost Increase (Problem 1)	30
Figure 7-4 - Computing the Partition Point on the Damage Axis (Problem 1)	31
Figure 7-5 - Computing the Conditional Expected Value (Problem 1)	33
Figure 9-1 - PDF of the Projected Cost Increase for Contractor B (Problem 2)	37
Figure 9-2 - PDF of the Projected Cost Increase for Contractor A (Problem 2)	38
Figure 9-3 - PDF for the Customer's Cost Increase (Problem 2)	38
Figure 9-4 - CDFs for Project Cost Increases for Contractors A and B and the Customer (Problem 2)	39
Figure 9-5 - Exceedance Probabilities for Project Cost Increases for Contractors A and B and the Customer (Problem 2)	39
Figure 9-6 - Conditional and Unconditional Expected Value of Project Cost Increases for Contractors A and B and the Customer (Problem 2)	40
Figure 10-1 - PDF for the Customer's Cost Increase (Problem 3)	45
Figure 10-2 - PDF for the Project Cost Increase for Contractor A (Problem 3)	46
Figure 10-3 - PDF for the Project Cost Increase for Contractor B (Problem 3)	45
Figure 10-4 - Comparative PDFs for Project Cost Increase for Customer and Contractors A and B (Problem 3)	48
Figure 10-5 - Example Triangular Distribution	48

CMU/SEI-93-TR-28 **v**

Figure 10-6 - Graphical Representation of the Results Summarized in Table 10-2	49
Figure 11-1 - Knowledge, Reality, and Perception of Uncertainties	53
Figure A-1 - SEI Taxonomy Structure	60
Figure A-2 - Taxonomy of Software Development Risks	6

Preface

The software development community has not been able to agree upon a set of measures to define the basic building blocks that can be used to generate cost and schedule estimates. For example, in most other engineering fields, cost estimates are based on basic measures; examples include BTUs, PSIs, length, height, width, weight, and throughput. In software, the measures may be years of experience, complexity of the requirements, software language to be used, and estimated number of lines of code, etc. The relationships among these factors and the cost or schedule estimate are not always clear, and this raises some questions as to the validity of the estimates in any particular case.

The following quotes excerpted from *Innovative Contracting Practices: Transportation Research Circular No. 386*, published by the National Research Council, December 1991 highlight the current dismal state of contracting practices:

- Innovative contracting techniques have been developed more in foreign countries than in the United States.
- Unfortunately, the lowest initial cost may not result in the lowest overall cost.
- In fact, current contracting practices provide little incentive for industry to be innovative.
- Agencies should develop contractor responsibility tests that reflect quality and performance factors; these tests should be examined and possible modifications should be developed.
- Indeed, the ability to assess quality and performance are directly related to the ability to assess risk.
- From a Summary of Questionnaire Findings:

This [pre-bid conferences] concept was the most popular, receiving a positive response from over 85 percent of the states participating in the survey. Better understanding of the scope of work, reduction in unanticipated construction conflicts, plan revisions, and other value engineering benefits can result from such conferences. Specialty jobs, especially fast-track projects, are most appropriate for this process.

- Risk management and assurance. End-result specifications and a determination of QA enter into this issue. Although not currently being practiced, many agencies are considering this concept for future application.

The questionnaire indicated that innovation has intensified in selected topic areas. Many agencies are implementing quality assurance-quality control (QA-QC) philosophies, contractor surveying, value engineering, off-peak time incentives, alternative bidding on structures, and other concepts. Additionally, many cost-saving and profitable concepts are being considered for future use and need to be developed further. On the other hand, many agencies expressed interest in receiving guidelines on other concepts that were not well understood.

Acknowledgements

We would like to thank Jim Lambert, David Gluch, Richard Schooff, Robert Kirkpatrick, Vijay Tulsiani, Archie Andrew, Richard Murphy, Scott Reed, Jack R. Ferguson, and Mike DeRiso for their valuable comments and suggestions. We also appreciate the technical editing assistance provided by Julia Deems. The conscientious retyping and continuously correcting new versions by Sharon Gingras of the University of Virginia and Leslie Chovan of SEI are greatly appreciated.

An Acquisition Process for the Management of Risks of Cost Overrun and Time Delay Associated with Software Development

Abstract: The ability to quantify risk is essential to the processes of budgeting and scheduling. During the process of hiring to complete specified tasks, customers must be able to verify contractor estimates and to make sound judgments on the risks of cost overruns and time delays. The following two questions are central to this paper: Do developers with little experience over-estimate or underestimate the complexity of the task because of their past experience, the assumptions they make, the models they select, and how they define the model parameters? What are the sources of risk associated with project cost estimation? How can such risk be quantified? To address these questions, this paper proposes a systematic acquisition process that is aimed at assessing and managing the risks of cost overruns and time delays associated with software development.

The proposed acquisition process, which is composed of four phases (listed below), is grounded on the following three basic premises: a) Any single-value estimate of cost or completion time is inadequate to capture and represent the variability and uncertainty associated with cost and schedule. Probabilistic quantification is advocated, using, in this paper, the fractile method and triangular distribution. b) The common expected value when used as a measure of risk, is inadequate; further, if used as the sole measure of risk, it may lead to inaccurate results. The conditional expected value of risk of extreme events is adopted to supplement and complement the common unconditional expected value. c) Probing the sources of risks and uncertainties associated with cost overruns and time delays in software development is essential for the ultimate management of technical and nontechnical risks. The Taxonomy-Based Questionnaire developed by the Software Engineering Institute is adopted.

These basic premises have led to the development of the following four phases in the proposed acquisition process: Phase I, constructing the probability density functions; Phase II, probing the sources of risks and uncertainties; Phase III, analyzing and comparing the significance and validity of the contractors' assumptions and premises regarding the likelihood of technical and nontechnical risks; and Phase IV, drawing conclusions on the basis of the accumulated evidence and ultimately selecting the contractors most likely to complete the project without major cost overruns or time delays. The three example problems are presented to demonstrate the construction of the probability density functions in Phase I and to explain in a more general way the effort involved in Phases II through IV.

1. Introduction

Three major classes of likely adverse consequences are prevalent in software development: risk of cost overrun, risk of time delay in the completion schedule, and risk of not meeting performance specifications. Here risk is defined as a measure of the probability and severity of adverse effects [Lowrance 1976]. The first two risks (cost overrun and time delay) are termed software nontechnical risks and the third (performance specifications) is termed software technical risk; more precise definitions can be found in Chittister and Haines [1993]. The focus of this paper is on the quantification (assessment) and management of software nontechnical risks, such as cost overruns and time delays.

The more central the role that software plays in overall system integration and coordination, the more likely the impact of delivery delay and/or of major cost overruns. Indeed, a series of auditing studies conducted by the General Accounting Office (e.g., GAO 1992) reveal an almost across-the-board epidemic of cost overruns and time delays in meeting completion schedules associated with software development for selected government-sponsored projects. A case in point is the C-17 airplane, cited in the previously mentioned GAO report, which experienced a major cost overrun and delivery delay.

Efforts have been made by some of the Source Selection Authorities (SSAs) and by their respective Boards in selecting contractors. Indeed, an SSA conducts a thorough search, examining, among other factors, the organizational capabilities of the contractor by evaluating performance history and in some cases a set of Key Practice Areas (KPAs), such as the processes of formal cost estimation and program management, as well as metrics for evaluating various performance criteria. Yet in spite of these efforts, the Department of Defense (DoD) has still had serious software delays.

The Software Engineering Institute (SEI) has also developed a methodology known as Software Capability Evaluation (SCE) [see for example, Humphrey & Sweet 1987] used to assess the software engineering capability of contractors. The SCE seeks to answer the question: Can the organization build the product correctly? It does so by considering three separate aspects of the contractor's expertise:

- organization and resource management,
- the software engineering process and its management, and
- available tools and technology.

Another tool, a risk taxonomy, also developed at SEI, addresses the sources of software technical risk and attempts to answer the question: "Is the organization building the right product?" [Carr et al. 1993]. These two processes (the SCE and the taxonomy), then, offer methods of *assessing* organizational processes and software technical risks; this paper presents, on the other hand, a process for *quantifying* the risks of project cost and schedule overruns. More is said on the risk taxonomy in Appendix A.

2. Overview of the Conceptual Framework

In this paper, we present a methodological framework for selecting a contractor that can assist the customer in minimizing the risks of project cost overruns and schedule delays. Although factors other than the selection of contractor(s) may decisively affect both software technical and nontechnical risks, they are treated here only as a general background; the interested reader is referred to Chittister and Haimes [1993, 1994] for a more in-depth discussion of these factors.

The process of selecting contractors is by itself quite complex; it is driven by legal, organizational, technical, financial, and other considerations—all of which serve as sources of risk. Because the world within which software engineering developed is non-deterministic, and because the central tendency measure of random events (i.e., the expected value of software nontechnical risk) conceals vital and critical information about these random events, special attention is focused on the variance of these events and on their extremes. Two approaches—the fractile method and triangular distribution—are adopted in this paper to quantify the probabilities of project cost overrun and delay in completion schedule. To capture the range of variation and the extremes of these probabilities, conditional expected values of extreme events are calculated using the partitioned multiobjective risk method (PMRM) [Asbeck & Haimes 1984] to supplement the common expected value of software nontechnical risk. To accomplish this objective, the fractile method, triangular distribution, and the PMRM are briefly introduced. Examples are included to clarify the appropriate application of these methods and to demonstrate their utility.

Figure 2-1 represents the conceptualization of the methodological framework. The framework can be viewed in terms of four major phases. The purpose of Phase I is to quantify the variances in the contractor's cost and schedule estimates by constructing probability density functions (PDFs) through triangular distributions, the fractile method, or through any other methods that seem suitable to the contractor. Extreme events are also assessed from these PDFs. In Phase II, using the SEI Taxonomy, interviews, and the PMRM, the sources of risks and uncertainties associated with each contractor are probed and evaluated; the assumptions and premises, which provide the basis for generating the variances in the contractor's estimates, are identified and evaluated; and the conditional expected value of risk of extreme cost overruns and time delays are constructed and evaluated. In Phase III, the significance, interpretation, and validity of each contractor's assumptions and premises are analyzed, ranked, filtered, and compared, and the probability of technical and nontechnical risks are assessed. In executing Phase III, three tools and methodologies are used: 1) an independent verification and validation team, 2) the risk ranking and filtering method, and 3) a comparative analysis. In the final phase, Phase IV,

conclusions are drawn on the basis of all the previously-generated evidence, including the opinions of expert judgment. The ultimate objective of the methodological approach is to minimize the following three objectives or indices of performance:

Minimize: { risk of project cost overrun
 { risk of project completion time delay
 { risk of not meeting performance criteria

Clearly, multiobjective tradeoff analysis, using, for example, the surrogate worth tradeoff (SWT) method, should be conducted where all costs and risks are kept and traded off in their own units. The objective of this paper is to develop scientifically-sound and pragmatic answers to some of the lingering problems and questions concerning the assessment and management of risks of those cost overruns and time delays associated with software engineering development.

It is constructive to discuss the four-phase acquisition process in more detail.

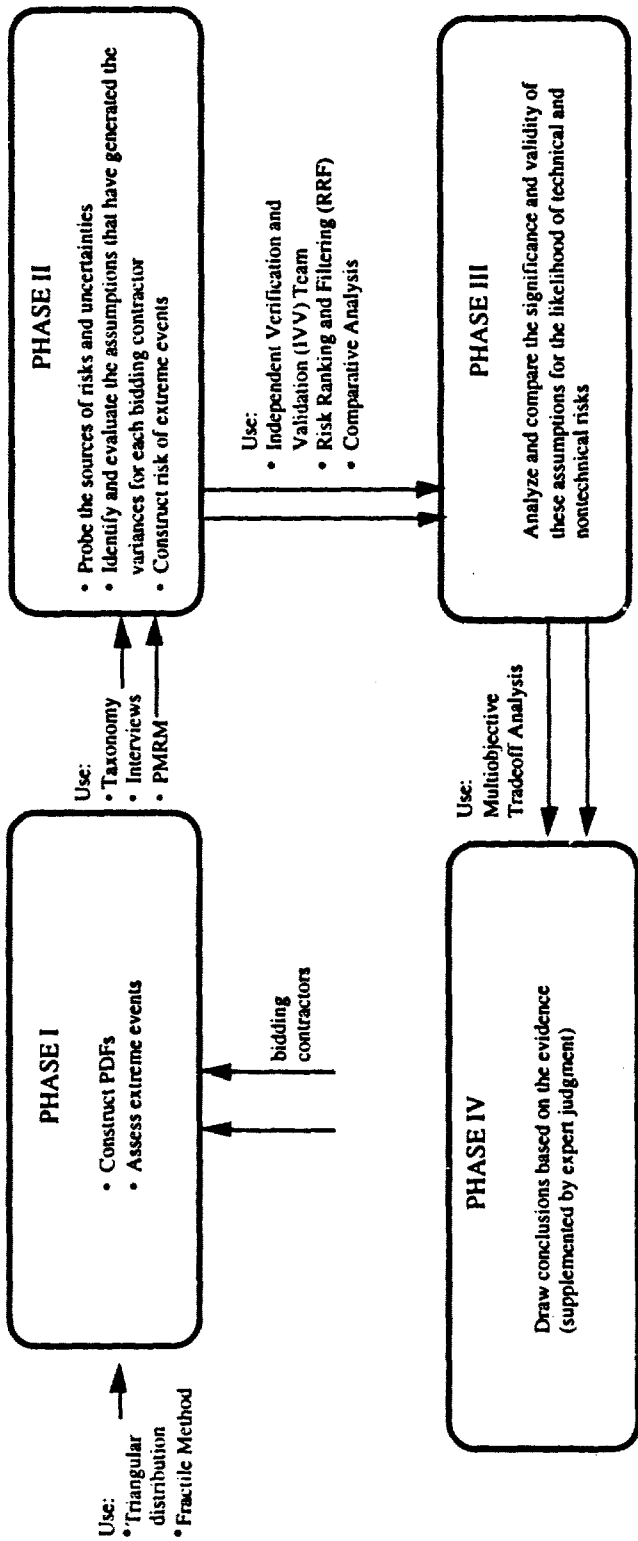


Figure 2-1 Proposed Acquisition Process

2.1. Phase I

Phase I will be demonstrated through the construction of the probability density functions (using the fractile method and triangular distribution) and through the assessment of extreme events (using the partitioned multiobjective risk method) by calculating the conditional expected value of extreme events to supplement the common unconditional expected value of cost overrun.

2.2. Phase II

Through the use of the Taxonomy-Based Questionnaire, interviews, and the quantification of risk of extreme events, Phase II provides a mechanism to probe the sources of risks and uncertainties; identify and evaluate the assumptions that have generated the variances for each bidding contractor; and construct the conditional expected value of risk of extreme events, $f_4(\bullet)$. The following discussion will focus on probing the sources of extreme events and the contractor's attitude:

2.2.1. Extreme Events

The shape of the probability density function, and particularly the behavior of the tail of the distribution, markedly influence the conditional expected value of extreme events. To demonstrate the effect of the tail of the distribution on projected cost overruns or time delays, all three examples have at least one project cost estimate with a long tail (i.e., a major cost overrun, albeit with a relatively low probability).

Most customers are primarily concerned with major cost overruns and time delays, even though they are also concerned with cost overruns and time delays of any magnitude. In other words, most customers want to prevent disastrous events that are beyond point β in Figure 6.1. The region to the left of β (e.g., cost overruns that would not exceed 10-15%) is commonly represented by the expected value measure of risk, $f_5(\bullet)$, whereas the region to the right of β is captured by the conditional expected value of risk of extreme events, $f_4(\bullet)$. It is with the help of the Taxonomy-Based Questionnaire that we can probe the sources of uncertainties and variabilities leading to $f_5(\bullet)$ and $f_4(\bullet)$. Indeed, the ultimate efficacy of risk assessment is its management through early identification, quantification, and *prevention*. Such a probe provides insights into the contractor's assumptions as to what can go wrong in a severe way that might cause the risk of extreme cost overrun or time delay to be catastrophic.

2.2.2. Contractor's Attitude

The Taxonomy-Based Questionnaire, along with the measurements of risk of cost overruns and time delays through f4(*) and f5(*), should explain not only the contractor's technical, financial, and other managerial assumptions and premises, but also the contractor's attitude toward risk. When a contractor's projection of lowest, most likely, and highest project costs falls, for example, in a close range, there are several possible explanations:

- The contractor is a risk seeker (a risk-averse contractor would have projected a much wider spread in the lowest, highest, and most likely project cost).
- The contractor is very knowledgeable and thus has confidence in the tight projections.
- The contractor is ignorant of the major technical details and complexity of the project's specifications; thus, major inherent uncertainties and variabilities associated with the project have been overlooked. Otherwise, the contractor would have projected a wider spread between the most likely and highest cost projections.

The Taxonomy not only constitutes an important instrument with which to discover the reasons for the uncertainties and variabilities associated with the contractor's projections; it also provides a mechanism that allows the customer to assess the validity and soundness of the contractor's assumptions. Indeed, the Taxonomy-Based Questionnaire, which is *systematic, structured, and repeatable* is an invaluable process with which the customer can elicit answers to the reasons for the contractors' variabilities. The accumulated assumptions about each contractor must then be compared and analyzed.

2.3. Phase III

In Phase III, an analysis and comparison of the significance and validity of the contractor's assumptions for the likelihood of technical and nontechnical risks are conducted. This is accomplished through the use of an Independent Verification and Validation (IVV) team, the Risk Ranking and Filtering (RRF) method, and other comparative analysis methods. In comparing assumptions, a number of issues may be addressed:

- stability of the requirements
- precedence of the requirements
- need for research about solutions

- politics and stability of funding
- overall knowledge and the lack thereof
- level of experience of key personnel
- maturity of technology
- maturity of the organization

In making these comparisons, the customer would be interested in ascertaining the reasons for the assumptions and determining whether they are based on knowledge or naivete, and whether the contractor has a conservative/risk-averse liberal/risk-seeking attitude. These issues will be highlighted in the example problems in subsequent discussions. The reader may consult, for example, the variances projected by Contractors A and B in Problem 3 (Figure 10-4). Contractor A is projecting a 50% cost overrun as the worst case, while contract B is projecting "only" a 40% cost overrun as the worst case. Is Contractor A more knowledgeable or more conservative than Contractor B? Or does the reason for this difference lie elsewhere? Is Contractor A risk-averse while Contractor B is risk-seeking? The information generated by the IVV team, the RRF method, and through other comparative analysis tools will be subjected to the expert judgment of the customer's team, leading to Phase IV of the proposed acquisition process.

2.4. Phase IV

Phase IV is the completion step where conclusions are drawn based on the accumulated evidence. Expert judgment is used in this phase in conjunction with multiobjective tradeoff analysis methods, such as the surrogate worth tradeoff (SWT) method (Haimes & Hall 1974). Adopting the systemic proposed acquisition process should markedly reduce the likelihood of major and catastrophic technical and nontechnical risks.

3. Critical Factors That Affect Software Nontechnical Risk

The proposed methodological framework for the quantification and management of software nontechnical risk—the risk of cost overrun and time delay associated with software development—is grounded on the premise that such management must be holistically-based. A holistic approach requires complete accounting of all important and relevant forces that drive the dynamics of cost overrun and time delay. Although a holistic view is advocated, introduced, and discussed in this paper, only limited aspects are ultimately quantified. Indeed, only a series of articles would do justice to the quantification of risks associated with all factors that embody software nontechnical risk. By their nature, the quantification and management of software nontechnical risk (and to a large extent software technical risk) embody:

- the customer (and the shadow client, i.e., the U.S. Congress, in the case of the Department of Defense),
- the contractor(s),
- the organizational interface between the customer and the contractor(s),
- the state of technology and knowhow,
- the complexity of the specification requirements,
- the add-on modifications and refinements,
- the availability of appropriate resources, and
- the models used for project cost estimation and schedule projection.

Since each element is in itself a complex entity with diverse dimensions, it is essential to recognize what characteristics of each component contribute to software nontechnical risk. Only by understanding the sources of risk can it ultimately be prevented and managed.

3.1. The Customer

The term customer is a misnomer because it connotes a singular entity. Yet, in most large-scale software engineering systems, such as DoD's systems, projects are initiated, advocated, nourished, and supported by multiple constituencies with some common, but often different, goals and objectives. Furthermore, for DoD projects, there is also the shadow customer—the U.S.

Congress which itself is influenced by various lobbyists, power brokers, and stakeholders. The existence and influence of this multiplicity of clients on the ultimate resources made available for the development of software engineering constitute a critical source of software risk. It is not uncommon for a pressure group to affect either the design specifications and/or the resources allocated for a specific DoD project and, thus, to impact on its final cost and its completion time. The "organizational maturity" level of the client is another factor that influences software nontechnical risk. A client that possesses internal capabilities to communicate with the contractor(s) on both the technical and nontechnical levels is more likely to have a better understanding and thus management of software nontechnical risk. This attribute will become more evident in this paper as specific quantitative information on the variances of cost and schedule is solicited in the proposed methodological framework.

3.2. The Contractor(s)

Elaborate procedures and protocols describing contractor selection for the development of software engineering have been designed and are being employed by government agencies and corporations. A commonly-accepted axiomatic premise is that the organizational maturity level of the contractor and the experience, expertise, and qualifications of its staff markedly impact the management of both software technical and nontechnical risks.

3.3. The Interface Between the Customer and the Contractor(s)

One of the dominant factors in initiating both technical and nontechnical risks can be traced to the organizational interface between the customer and the contractor(s). Adequate and appropriate communication between the two parties, and the understanding and the appreciation of each other's role throughout the life cycle of the software development process are imperatives to the prevention and/or control of potential risks.

3.4. The State of Technology and Knowhow

Although many consider the contractor's access to knowhow and the access to appropriate technology to be major factors in controlling software technical risk, they also impact software nontechnical risk. In particular, the lack of access to appropriate technology or a deficient knowhow of the contractor's staff is likely to cause a measurable time delay in the completion of a project and is also likely to cause cost overrun.

3.5. The Complexity of the Specification Requirements

The more unprecedented the client's project specifications in terms of advanced and emerging technology, the higher the risk of time delay in its completion and of its cost overrun. Most systems developed by the DoD are advancing the state of the art in some field of technology, e.g., software development, stealth, propulsion, and satellites. The requirements in these fields are necessarily complex since the parameters are constrained by the task and are frequently subject to modifications because of changing technology.

3.6. The Add-On Modifications and Refinements

Add-on modifications and refinements are viewed by many as an Achilles' heel in terms of software nontechnical risk. Although these add-on modifications are often associated with software nontechnical risk, they also constitute a critical source of software technical risk. This is because not all modifications are appropriately related to and checked against the original design to ensure ultimate compatibility and harmony. Very large and complex systems are difficult to manage. Systems are now developed by multiple companies, each having its own area of expertise, and changes often ripple through the entire system. A wide range of factors may cause mid-course modifications; however, three categories of causes emerge:

- a) **Threat or Need Change:** When a new threat is projected or a new need is contemplated.
- b) **Improved New Technology:** When a new technology provides improved performance or quality, such as a new sensor.
- c) **Replacing Obsolete Technology:** When the pre-selected technology becomes obsolete before the contract has even begun or completed.

3.7. The Availability of Appropriate Resources

One open secret in government procurement and occasionally in the private sector is the level of pre-allocated funds for a specific project. The competitive zeal of contractors often outweighs the technical judgment of their professional staff; the outcome is a bid that is close to the pre-allocated funds by the client even though it is clear to the bidder that the job with its specification requirements cannot be delivered at that level of funding. This not-uncommon phenomenon is

dramatically illustrated by numerous documented examples by Hedrick Smith [1988] in his book *The Power Game: How Washington Works*:

The standard technique is to get a project started by having the prime contractor give a low initial cost estimate to make it seem affordable and wait to add fancy electronics and other gadgets much later through engineering "change orders," which jack up the price and the profits. Anyone who has been through building or remodeling a house knows the problem. "This is called the buy-in game," an experienced Senate defense staff specialist confided.

3.8. The Models Used for Project Cost Estimation and Schedule Projection

A number of models are used to estimate project cost and completion schedule. Constructive Cost Model (COCOMO) [Boehm 1981] and Program Evaluation and Review Technique (PERT) are representative examples. Models can be potent tools when they are well understood, are supported by an appropriate data base, and adhere to the assumptions upon which they are designed to operate. The complexities of such models, however, often result in their misuse and/or invalid interpretations of their results. They thus ironically become a source of software nontechnical risk. The successful application of the proposed methodological framework, however, does not depend on the specific model used by either the contractor or the customer to estimate either the cost or the schedule.

From the above it seems that the sources that contribute to software nontechnical risk are organizational and technical in nature; they stem from failures associated with the contractor as well as the customer. In terms of the contractor, these failures primarily originate from and are functions of such elements as:

- the organizational maturity level
- the process and procedures followed in the assessment of the project's cost and schedule
- the level of honesty exhibited by management in communicating the real cost and schedule to the customer (and of course vice versa)
- the extent and level of new and unprecedented technology imposed on the project
- the level of experience and expertise of the staff engineers in software engineering in general and in the application domain in particular

- the level of experience and expertise of the management team in software engineering
- the overall competence of the team developing the software
- financial and competitive considerations
- immature technology, methods, and tools
- using technology in new domains
- combining methods and tools in new ways and using them in a new software development environment
- requirement modifications causing changes in the system's architecture

In terms of the customer, the nature of organizational failures partially overlap those of the contractor's, but also have distinctive characteristics:

- the process and procedures followed in the assessment of the project cost and schedule
- the level of specificity at which the system and software requirements are detailed
- the number of changes and modifications requested by the customer during the software development process. These changes (which generally introduce many new errors) are often not harmonious with earlier specification requirements.
- the commitment of project management (associated with the customer's organization) to closely monitor and oversee the software development process
- the specific requirements for technology, e.g., specific compiler and data base management systems
- the level of honesty exhibited by management in communicating the real cost to the "real client" (e.g., the Department of Defense as a client and the U.S. Congress as the "real client")

4. Basis for Variances in Cost Estimation

Most, if not all, developers of large complex software systems use cost models to estimate their costs. These models are based on a set of relationships based on such parameters as the size and complexity of the software, the experience level of the software developer, and the type of application within which the software will be used. Different models generate different weights or levels of importance for these parameters, and not all models use the same parameters. Therefore, one cost model can lead to a radically different cost estimate than another just on the basis of which parameters are used in the model and how they are implemented. Even if the parameters are used consistently, however, different developers will probably not agree on the value or weight of the parameters in the first place. Many organizations, in fact, consider their interpretations of these parameters to contribute to their "competitive edge" because the definition affects their ability to determine costs accurately. For example, an organization that has experience in developing Space system software may not have the same perception of difficulty when developing a complex avionic software system as would an organization that has significant experience in that area. Their understanding of Space systems, however, will alter their definition of the avionic system parameters. Do developers with little experience overestimate or underestimate the complexity of the task because of how they define these parameters? As stated in the beginning, these are questions central to this paper: What are the sources of risk associated with project cost estimation? How can such risk be quantified?

Although creating, maintaining, and updating project cost estimation metrics and parameters are extremely important for an organization, it is nevertheless unlikely that a future project will be similar enough to previous projects to merit directly importing these metrics or parameters; such metrics and parameters may not be directly applicable without appropriate modifications. Indeed, cost estimators are required to (and do) use judgment when applying these parameters to a new project requirement. Furthermore, cost estimation constitutes a critical area with regard to the sources of risk for software development, which is without parallel to other fields. For example, if a contractor were estimating the cost to construct a building with 50 stories, yet the contractor had previously only built structures with a maximum of 10 stories, then the contractor would not just increase the estimate five-fold. In fact, the contractor may question the basic foundations and relevance of extending the 10-story model to the new structure parameters. However, in software, it is not uncommon to increase estimates for new projects by a factor of five from previous projects of one-fifth the size and complexity. Many new systems have size estimates of over 1,000,000 lines of code even though the developers have little experience with systems of this size.

Another example is in the use of commercial-off-the-shelf (COTS) software. The original assumption that a commercial database management system (DBMS) can be used to meet customer requirements may change if the customer requires features not supported by DBMS suppliers. Such changes may have serious ramifications for the cost estimate depending on how the developer plans to solve the problem. If the developer chooses to subcontract out the effort and deal with the subcontractor as he does the DBMS vendor, this has ramifications for the risk associated with the subcontractor--an important subject that will be discussed later. The alternative is for the developer to undertake the development of his or her own DBMS. This requires an additional set of assumptions, design parameters, and judgments regarding the architecture, size, experience level, domain knowledge, software engineering knowledge, and the support environment needed to develop the DBMS. Each of these assumptions, parameters, and judgments has some uncertainty associated with it. This uncertainty contributes to the overall risk in the cost estimate. If the developer chooses to subcontract the DBMS development to an outside vendor, then the issue for the contractor is understanding and accounting for the set of assumptions that are made by the subcontractors on the DBMS and on the system architecture.

The ability of the developer to make valid assumptions and design decisions is usually based on a set of metrics; these metrics can either be based on current measurements or on past performance. Either way, however, there has to be an agreed-upon set of measures that is being evaluated (such as the number of lines of code needed to accomplish specified tasks or productivity rates in terms of lines of code per hour). The difficulty with software development is that the community has not agreed upon basic measures, such as how to count lines of code or how to measure productivity. Also the difficulty with using performance history is that the systems under development are sufficiently different such that history may not adequately reflect the new parameters accurately.

5. The Quantification of Software Nontechnical Risk and the Evaluation of Variances

The premise of this paper is that the manner in which the customer selects a contractor affects the risk of cost overruns, time delays, and failure to meet performance specifications. Therefore, the proposed methodological approach requires that the contractor provide the customer with more than fixed, deterministic values for the cost and time schedule to deliver software with prespecified performance requirements.

Building on quantitative probabilistic assessment, the contractor is asked to submit either an estimated variance of the cost and time to completion, or a probability density function (See, for example, Figure 4.) for the projected project cost and time schedule. (A similar approach can be adopted to quantify software technical risk.) This variance may be generated from, for example, a triangular distribution, where the contractor specifies the lowest possible cost, the highest possible cost, and the most likely cost of the project. Alternatively, the contractor may choose to provide the variance estimate through the fractile method. (The construction of a PDF using a triangular distribution and the fractile method will be subsequently discussed.) Similar estimates are to be provided for the completion schedule. *The information provided by the contractor constitutes the basis for the generation of probability density functions (PDFs) associated with the project's cost and its completion time.* Assuming that the customer (with the assistance of a MITRE-type organization, if needed) is also capable of providing (for comparative purposes with the clients' variances) basic variance information, which allows the generation of the customer's PDFs for the projected cost and completion time, then the following method will be useful. A mechanism is developed here that enables the customer to compare various contractors' probabilistic estimates of several attributes and characteristics to its own estimate. To use either of the two approaches to estimate cost or schedule variances, the contractor should familiarize the team making such estimation with the intricacy of these approaches and alert it to the cognitive biases inherent in such an estimation process. In their quest to quantify these human biases, Alpert and Raiffa [1982] conducted several experiments over two decades ago and arrived at the conclusion that with appropriate training, the use of the fractile method can be very effective. The question of gaming and the manipulation of the approach by some contractors to gain advantage will be discussed in Section 11.

The proposed methodological framework requires a number of steps. First, the customer requires that each bidding contractor submit either basic information on the cost and completion time variances or the corresponding PDFs. In the latter case, the contractor may choose to use a

triangular distribution, the fractile method, or any other means that the contractor believes will provide the most accurate estimate. Clearly, the contractor can and is likely to use models and other tools to generate the required PDFs. At the same time the customer's staff will generate its own PDFs for cost and completion time. The customer is now able to compare not only the expected values (the means) of each contractor's cost and completion time, but also the variances of these estimates. Furthermore, a comparison of the extremes of each PDF provides valuable information to the customer about each contractor's capabilities and compatibility. Although some of the efficacious attributes of the methodological framework will be better understood after we introduce the section on Risk of Extreme Events and the example problems, an overview of the required steps may clarify the process:

1. Use the fractile method, the triangular distribution, or any other approach to quantify the variances associated with project cost estimates and completion schedule.
2. Assess the contractor's capability to deliver the product and to estimate the likely variance of the project's cost and schedule through SEI's Software Risk Taxonomy-Based Questionnaire. The SEI taxonomy and the accompanying questionnaire provide a framework for identifying the technical uncertainties in a project and the root causes for these uncertainties. Also, it provides a method for assessing the honesty and credibility of the contractor's analysis and figures.
3. Evaluate, in a quantitative way, any discrepancy between the variance assessments of the contractor and the customer. (The quantification is likely to lead to significant information about the likelihood of extreme events and their potential consequences on the entire project.)
4. Investigate and understand the contractor's assumptions in estimating variances. This information will enable the customer's staff to take appropriate measures to mitigate software nontechnical risk.
5. Integrate the information on the contractor derived from (a) the quantitative variances received on the projected cost and completion schedule with (b) the results generated from the Software Risk Taxonomy-Based Questionnaire.
6. Use the risk ranking and filtering (RRF) method to rank the risks associated with each prospective contracting organization; then, compare those risks against one another and against an established norm.

Although these methods and processes may not provide an optimal approach for selecting the best or most valid estimate, they do provide a foundation that is systematic and repeatable to allow the evaluators to gain significant knowledge and insight into the estimators' assumptions. This insight is critical from two perspectives; it enables the customer to:

- evaluate whether the contractors' estimates and assumptions are valid and consistent with the specifications, and
- establish a foundation by which to evaluate and judge future changes to cost and schedule estimates.

These methods and processes also provide a mechanism for the customer's evaluation team to document the assumptions and risks in a cost or schedule estimate, identify the root issues associated with these assumptions and risks, and organize this information within the taxonomy framework. This information can then be used to measure progress and can also be used as a metric against future cost and schedule estimates.

6. Risk of Extreme Events

In general, the estimates of the most likely project cost provided by the dominant number of contractors will be within a close range of one another. Since assessing and ultimately preventing potential major cost overruns and time delays are of major concern in this paper, our interest here is in what can go wrong in the extremes, i.e., in the behavior of the tail of the distribution. This can be captured best through the conditional expected value of extreme events. The conditional expected value of risk, denoted by $f_4(\cdot)$ (which will be defined later), can provide valuable information that supplements and complements the average cost or most likely cost estimates.

Most analysts, who use probabilistic quantitative methods to measure the risk of project cost overruns and delays in its completion schedule, resort to the most common mathematical construct for the quantification of risk--the expected value of risk. Whether the probabilities associated with the universe of events are viewed by the analyst as discrete or continuous, the expected value of risk is an operation that essentially multiplies each event by its probability of occurrence and sums (or integrates) all these products over the entire universe of events. This operation literally commensurates adverse events of high consequences and low probabilities of exceedance with events of low consequences and high probabilities of exceedance. Indeed, the expected value masks the extremes and hides the effects of less likely outcomes.

The misuse, misinterpretation, and fallacy of the expected value when it is used as the sole criterion for risk in decisionmaking are discussed elsewhere [Haimes 1993, Asbeck & Haimes 1984]. Many experts are becoming convinced of the grave limitations of the traditional and commonly used expected-value concept and so are augmenting the expected value of risk with a supplementary measure--the conditional expectation--by which decisions about extreme and catastrophic events are *not* averaged out with more commonly occurring high-frequency/low-consequence events.

The partitioned multiobjective risk method (PMRM) is a risk analysis method developed for solving probabilistic multiobjective problems with a focus on extreme events [Asbeck & Haimes 1984]. Instead of using the traditional expected value of risk, the PMRM generates a number of conditional expected-value functions, known as risk functions, which represent the risk given that the damage falls within specific ranges of the probability of exceedance or within a range of adverse consequences (generically termed as damages). Before the PMRM was developed, problems with at least one random variable were solved by computing and minimizing the unconditional expectation of the random variable representing the specific damage. In contrast,

the PMRM isolates a number of damage ranges (by specifying partitioning probabilities) and generates conditional expectations of damage given that the damage falls within a particular range. In this manner, the PMRM can generate a number of risk functions, one for each range, which are then augmented with the original optimization problem as new objective functions. In this paper the discussion will be limited to one conditional expected value of extreme events, denoted by $f_4(\bullet)$.

The conditional expectations of a problem are found by partitioning the problem's probability axis and mapping these partitions onto the damage axis. The damage axis in this case can be project cost overrun in terms of dollars or percentage of overage above the contracted level; alternatively, the damage can represent a time delay either in terms of months or weeks or in terms of percentages in relation to the original time schedule. Consequently, the damage axis is partitioned into corresponding ranges. *A conditional expectation is defined as the expected value of a random variable given that this value lies within some prespecified probability range (or within some prespecified damage range).* Clearly, the values of conditional expectations are dependent on where the probability axis (or the damage axis) is partitioned. The choice of where to partition is made subjectively by the analyst in response to the extreme characteristics of the decision-making problem.

A continuous random variable X of damages (e.g., cost overrun or time delay) has a cumulative distribution function (CDF) $P(x)$ and a probability density function (PDF) $p(x)$, which are defined by the relationships

$$\text{CDF: } P(x) = \text{prob}[X \leq x] \quad (1)$$

and

$$\text{PDF: } p(x) = \frac{dP(x)}{dx} \quad (2)$$

The CDF represents the *nonexceedance probability* of x . The *exceedance probability* of x is defined as the probability that X is observed to be greater than x and is equal to one minus the CDF evaluated at x .

The expected value, average, or mean value of the random variable X is defined as

$$E[X] = \int_0^{\infty} x p(x) dx \quad (3)$$

For the purpose of this paper, a modified version of the PMRM is presented to simplify the mathematical discussion and to focus the analysis on the conditional risk of extreme events. Let $1 - \alpha$, where $0 < \alpha < 1$, denote an exceedance probability that partitions the domain of X into two ranges. On a plot of exceedance probability, there is a unique damage β on the damage axis that corresponds to the exceedance probability $1 - \alpha$ on the probability axis. Damages (e.g., cost overruns or time delays in project completion schedule) less than β are considered to be of low to moderate severity; damages greater than β are of high severity. The partitioning of risk into two severity ranges is illustrated in Fig. 6-1. If the partitioning probability α is specified, for example, to be 0.95, then β is the 95th percentile.

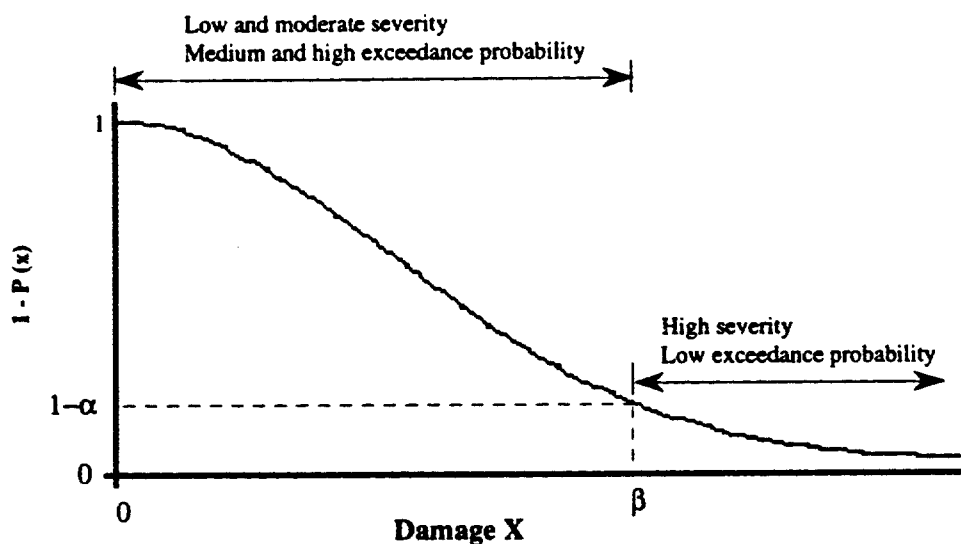


Figure 6-1 Mapping of the Probability Partitioning onto the Damage Axis

The conditional expected damage (given that the damage is within that particular range) provides a measure of the risk associated with the range. The measure of conditional expected value of risk of interest here is that of low exceedance probability and high severity, denoted by $f_4(\bullet)$. High severity may mean a high cost overrun or a high time delay in the project's scheduled completion. The function $f_4(\bullet)$ is the expected value of X , given that x is greater than or equal to β :

$$f_4(\bullet) = E[X | x \geq \beta] \tag{4}$$

For any probability of exceedance, one can generate the traditional, unconditional (common) expected value of risk (of cost overrun and/or of time delay) denoted by $f_5(\bullet)$, and the conditional expected value of risk of extreme events (of same) denoted by $f_4(\bullet)$. Note that

$$f_4(\bullet) = \frac{\int_{\beta}^{\infty} x p(x) dx}{\int_{\beta}^{\infty} p(x) dx} \quad (5)$$

$$f_5(\bullet) = \int_0^{\infty} x p(x) dx \quad (6)$$

where $p(x)$ is the probability density function.

The use of Equations 5 and 6, which are shown above, will be discussed in detail in subsequent sections through three sample problems. The three sample problems have been constructed in such a way so as to demonstrate the importance of the Taxonomy-Based Questionnaire along with the two measurements of risk (the unconditional expected value of risk, $f_5(\bullet)$, and the conditional risk of extreme events, $f_4(\bullet)$) in understanding the contractor's premises and attitudes.

Problem 1 demonstrates the use of the fractile method in the construction of a probability density function for one contractor.

Problem 2 extends the discussion in the previous example by including two contractors and a customer, using the fractile method. To focus on the methodological approach, the expected values and the conditional expected values of cost increase for all three parties are given in percentages. For comparative purposes, one needs to evaluate costs (estimated by each party) in addition to the variations in the percentages as is the case in this example problem.

Problem 3 demonstrates the use of the triangular distribution with one customer and two contractors. Here again, for pedagogical purposes only the percentages in the projected cost increases are addressed.

7. Problem 1

The Department of Defense (DoD) is considering the introduction of a new strategic airplane that will constitute the flagship of the Air Force as we enter the third millennium. Aware of the powershift from hardware to software in technology and the emerging centrality of software as the overall system integrator and coordinator, DoD considers the development of software for this airplane to be of paramount importance [Chittister & Haines 1994]. The Air Force commissions the assistance of a support organization to develop, in collaboration with its own staff, specifications and a request for proposal (RFP) for designing, prototyping, and developing the software needed for the flagship airplane. Following a detailed and tedious process of qualifying prospective bidders, the Air Force issues an RFP for the development of the required software engineering. This time, however, the RFP includes items that had not been requested previously. For example, the RFP requires that each contractor provide variances along with the estimated project's cost and completion schedule, instead of the commonly-practiced requirement of single deterministic values. The RFP leaves it up to the contractors to determine the form that these variances take, including, if the contractor so desires, the type of PDF selected for each estimate. The Air Force and its support team, planning to use the same approach themselves in evaluating the various proposals, recommends in the RFP the optional use of the fractile method or the triangular distribution when complete statistical information is not readily available.

To capture the mathematical details entailed in the process of developing representative PDFs for cost and completion time, a step-by-step procedure using the fractile method (adopted by Contractor A) is presented here. The team from Contractor A estimates a most likely cost of \$150 million. After considerable brainstorming sessions, the following information emerges:

- Best case project cost increase = 0% (i.e., project cost is \$150 million)
- Worst case project cost increase = 50% (i.e., project cost increase is \$75 million, for a total of \$225 million)
- Median value of project cost increase (equal likelihood of being greater or less than this value) = 15% (i.e., project cost increase is \$22.5 million, for a total of \$172.5 million)
- 50-50 chance that the actual project cost would be within 5% of the 15% median estimate (i.e., project cost increase is $(15 \pm 5)\%$)

From the above information, the following fractiles (percentiles) are readily determined.

- The best scenario of no cost overrun (0% cost increase, i.e., a total cost of \$150 million) represents the 0.00 fractile (0th percentile).
- The worst scenario of 50% cost overrun (a total cost of \$225 million) represents the 1.00 fractile (100th percentile).
- The median value of 15% cost overrun (a total cost of \$172.5 million) represents the 0.50 fractile (50th percentile).
- The 0.25 fractile (25th percentile) is $(15-5)\% = 10\%$ increase over \$150 million (a total cost of \$165 million).
- The 0.75 fractile (75th percentile) is $(15+5)\% = 20\%$ increase over \$150 million (a total cost of \$180 million).

The above assessment of project cost is summarized in Table 7-1 and is used as a basis for constructing the corresponding cumulative distribution function (CDF). (See Figure 7-1.)

Table 7-1 Tabular CDF for Problem 1

Fractile	Project Cost Increase (%)	Project Cost (\$ million)
0.00	0	150
0.25	10	165
0.50	15	172.5
0.75	20	180
1.00	50	225

The CDF (Figure 7-1) can now be represented in terms of a PDF (Figure 7-2). To construct the PDF, one must be guided by the following principles:

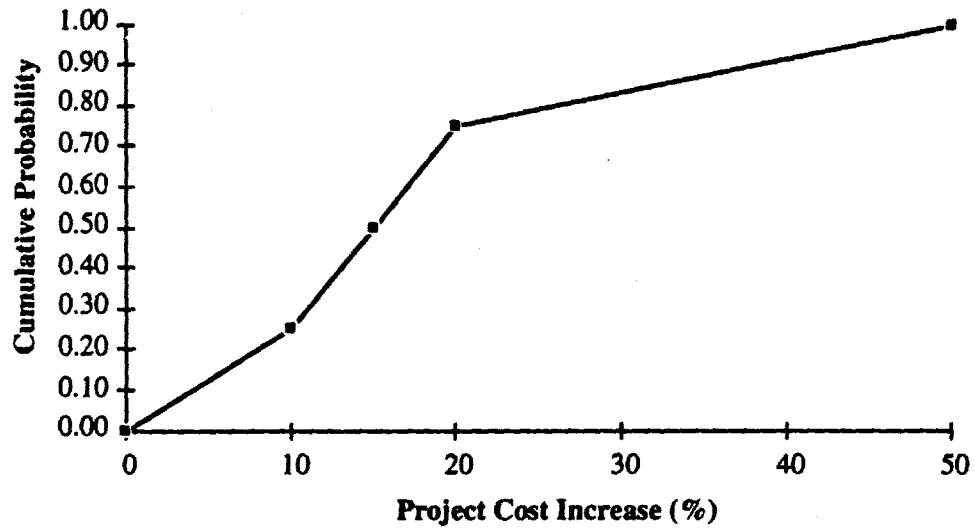


Figure 7-1 Graphical CDF for Project Cost Increase (Problem 1)

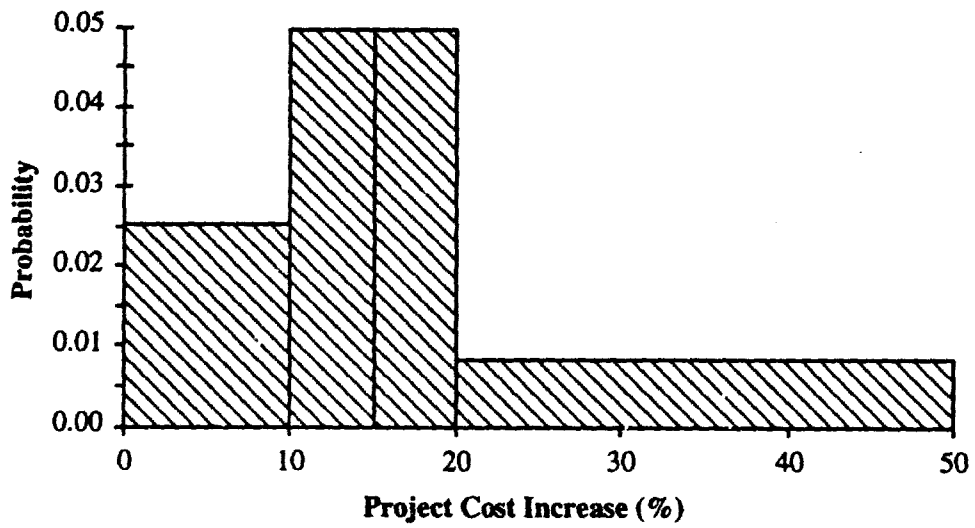


Figure 7-2 Probability Density Function for Project Cost Increase (Problem 1)

- 1) The area under the shaded area (the PDF) must be equal to 1.
- 2) The first quartile in Figure 7-1 (representing 25% of the probabilities) spans a cost overrun from 0% to 10%.

Thus, the corresponding area of the PDF (Figure 7-2) must be equal to one-fourth of the total area, i.e., 0.25. Dividing 0.25 by 10 yields a height of 0.025 for the first rectangle in Figure 7-2. Similarly for the other three quartiles, each of the second and third quartiles spans 5% of project cost increase. Thus, the height of the rectangle of the PDF (Figure 7-2) is 0.25 on the probability axis and when divided by 5 yields a height of 0.05 on the probability axis. Finally, the last quartile spans a cost overrun of 30% (from 20% to 50%). Thus, the height of the rectangle (on the probability axis) is 0.25 divided by 30 which yields a height of 0.008. Figure 7-3 depicts the exceedance probability (I-CDF) vs. project cost increase. To focus on the exceedance probability of a major cost overrun, say between 20% and 50%, only that part of Figure 7-3 is depicted in Figure 7-4. Note that by just using basic rules from geometry, one can relate the exceedance probability to any project cost increase x , for $20\% \leq x \leq 50\%$.

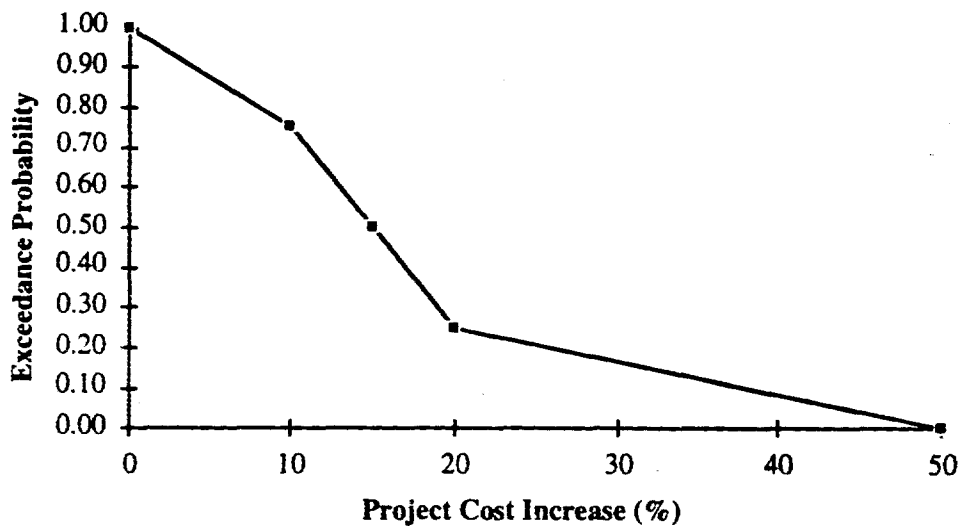


Figure 7-3 Exceedance Probability for Project Cost Increase (Problem 1)

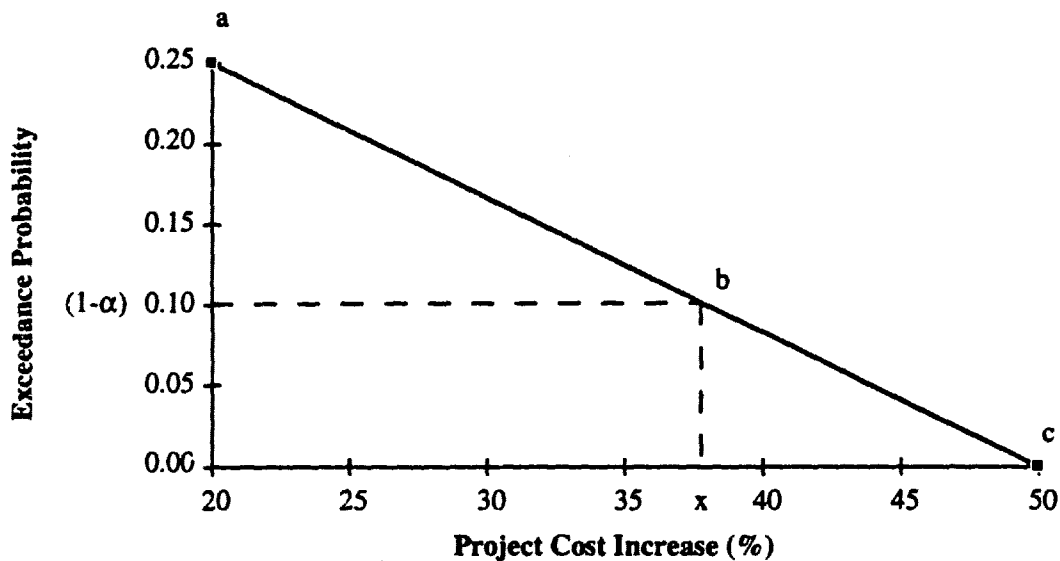


Figure 7-4 Computing the Partition Point on the Damage Axis (Problem 1)

The expected value of the percentage of project cost increase can be determined from geometry (Figure 7-2):

$$\begin{aligned}
 f_5(\bullet) &= 0.25 \left[0 + \frac{(10 - 0)}{2} \right] + 0.25 \left[10 + \frac{(15 - 10)}{2} \right] + 0.25 \left[15 + \frac{(20 - 15)}{2} \right] \\
 &\quad + 0.25 \left[20 + \frac{(50 - 20)}{2} \right] \\
 &= 0.25 (5) + 0.25 (12.5) + 0.25 (17.5) + 0.25 (35) \\
 &= 0.25 (70) = 17.5\% \text{ (i.e., total cost of } \$ (150 + 26.25) \text{ million)}
 \end{aligned}$$

The expected value of the percentage of project cost increase may also be calculated using Equation 3. This is given in Appendix B.

Note that the expected value of cost overrun of \$26.25 million (i.e., total cost of \$176.25 million) does not provide any vital information on the probable extreme behavior of project cost. Also note that there is a one-to-one functional relationship between 0.1 probability of exceedance and 38% cost overrun; this relationship is depicted in Figure 7-4 and is generated as follows: (Here we are interested in the probability of exceedance of 0.1, i.e., $\alpha = 0.90$, or $(1-\alpha) = 0.10$.)

$$\frac{x - 20}{50 - 20} = \frac{ab}{ac} = \frac{0.25 - (1 - \alpha)}{0.25}$$

$$\text{Thus, } x = 30 - \frac{30(1 - \alpha)}{0.25} + 20$$

$$= 38\% \text{ for } \alpha = 0.9$$

Alternatively, we can compute from Figure 7-4 the partition point x (the percentage of increase in cost) that corresponds to a probability of 0.1 as shown below:

$$(1 - \alpha) = (50 - x) h$$

where h is the height in the probability axis

$$h = \frac{0.25}{50 - 20} = 0.0083$$

$$x = 50 - \left(\frac{1 - \alpha}{h} \right)$$

$$= 50 - \frac{(1 - 0.9)}{0.0083}$$

$$= 38\% \text{ for } \alpha = 0.9$$

The conditional expected value of project cost can be calculated for a couple of scenarios to shed light on the behavior of the tail of the PDF. For example, given (from Figure 7-4) that there is 0.1 probability of project cost overrun that would exceed 38% of its original scheduled budget, management might be interested in answering the following question: What is the conditional expected value of extreme cost overrun beyond the 38% (or extreme cost overrun with exceedance probability that is below 0.1)? Or posed differently, within the range of exceedance probabilities between 0.1 and 0.0 and range of cost overruns between 38% and 50%, what is the expected value of project cost overrun? Note that 1) the maximum cost overrun was predicted not to exceed 50%, 2) the conditional expected value is the common expected value limited between specific levels of cost overruns instead of the entire range of possible cost overruns, and 3) the expected value is a weighted average of possible cost overruns multiplied by their corresponding probabilities of occurrence and summed over that entire range.

Using Equation 3, the common, unconditional expected value of cost overrun, $f_5(\bullet)$ was calculated several pages earlier to be 17.5%. Similarly, the conditional expected value of cost overrun under the scenario of 0.1 probability of exceeding the original cost estimate (by 38% or by \$57 million) computed using equation 5 yields $f_4(\bullet) = 44\%$. (See Appendix C.) Note that the PDF of cost

overrun portion from 20% and beyond is a linear function kx . Alternatively, the conditional expected value can be computed as the mean of the shaded area in Figure 7-5.

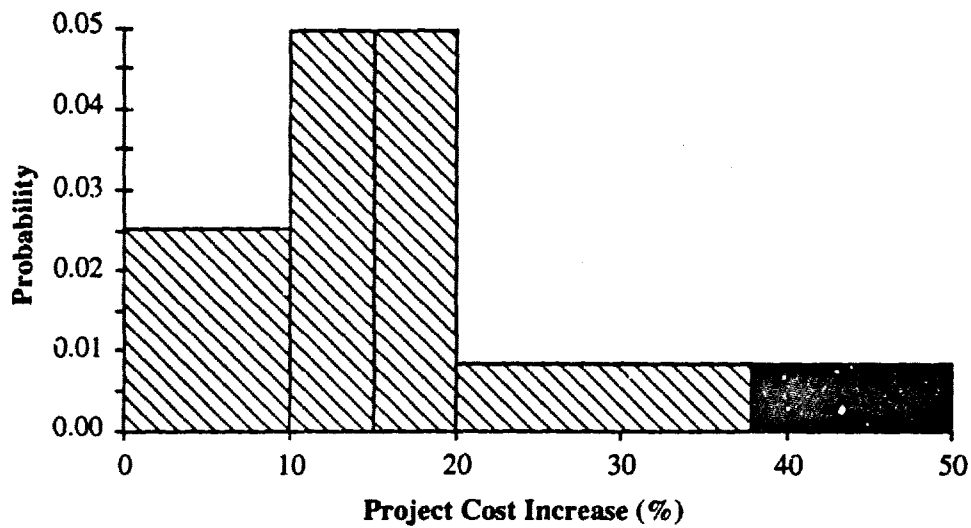


Figure 7-5 Computing the Conditional Expected Value (Problem 1)

$$f_4(e) = 38 + \frac{50 - 38}{2} = 44\%$$

In other words, the adjusted expected value of cost overrun, when it is in the range of 38% to 50% of the original scheduled cost is 44%.

8. General Notation

Unless the project is a cost-plus contract, the interpretation of these results should alarm top management of Contractor A: although the expected cost overrun of the proposed budget is 17.50% above the budgeted cost of \$150 million, there is a 10% chance that the cost overrun will exceed 38% of the budgeted cost! Furthermore, at a 10% chance of cost overrun, the conditional expected value of cost overrun that exceeds 38% is 44% above the original budget, i.e., an exceedance of \$66 million; in other words, under these conditions, the expected value of the total cost will be $(\$150 + 66)$ million = \$216 million.

It is worthwhile to clarify at this point the meaning of the two terms of cost overrun: 38% and 44%. The 38% cost overrun corresponds to a single probability point and is derived directly from Figure 7-4. The 44% cost overrun, on the other hand, represents an expected value, the averaging of all the probabilities from 0.10 to zero multiplied by the corresponding cost overruns from 38% to infinity, summed as appropriate and scaled. Thus

$$f_4(\bullet) = E [X | \geq 38\% \text{ cost overrun}] = 44\%$$

or

$$f_4(\bullet) = E [X | \geq \$207 \text{ million}] = \$216 \text{ million}$$

Similar analysis using a triangular distribution is introduced in a subsequent section.

To make the most use of the contractors' cost and completion schedule variances as represented by the PDFs, the Air Force develops the customer's version of these PDFs.

To streamline the discussion, the following symbols are used:

Let	$p_i(x;c)$	=	the PDF for the i^{th} project cost
	$p_i(x;t)$	=	the PDF for the i^{th} project completion time
	$P_i(x;c)$	=	the CDF for the i^{th} project cost
	$P_i(x;t)$	=	the CDF for the i^{th} project completion time
	α	=	the partitioning probability

where $i = 0$ denotes the customer;

$i = 1, 2, 3, \dots, n$ denote the i^{th} contractor;

c denotes project cost, and

t denotes completion time.

$f_4(\alpha; i, c)$ = the conditional expected value of cost ($i = 0, 1, 2, \dots, n$) at a partitioning probability α . As noted, the index $i = 0$ is reserved to denote the customer's estimate.

$f_4(\alpha; i, t)$ = the conditional expected value of completion time ($i = 0, 1, 2, \dots, n$) at a partitioning probability α

$f_5(i, c)$ = the conventional (unconditional) expected value of cost ($i = 0, 1, 2, \dots, n$)

$f_5(i, t)$ = the conventional (unconditional) expected value of completion time ($i = 0, 1, 2, \dots, n$)

Table 8-1 Summary of Available Information

Symbol	Description
$p_i(x; c)$	Probability density function (PDF) for project cost
$p_i(x; t)$	Probability density function (PDF) for project completion time
$P_i(x; c)$	Cumulative distribution function (CDF) for project cost
$P_i(x; t)$	Cumulative distribution function (CDF) for project completion time
$f_5(i, c)$	Conventional (unconditional) expected value of cost
$f_5(i, t)$	Conventional (unconditional) expected value of completion time
$f_4(\alpha; i, c)$	Conditional expected value of cost at partitioning probability α
$f_4(\alpha; i, t)$	Conditional expected value of completion time at partitioning probability α

$i=0$ denotes customer
 $i=1, 2, 3, \dots, n$ denote contractors

9. Problem 2

This example extends the previous problem by adding another contractor to the bidding process and by including the customer's estimates as well. Table 9-1 summarizes the estimated fractiles for the customer and for Contractors A and B. Table 9-2 summarizes the unconditional expected value of cost overrun for the customer's estimate and for the two contractors. Table 9-2 also lists the conditional expected values of cost overrun for four different partitioning points. Figures 9-1, 9-2, and 9-3 depict the PDFs of the projected cost variances for Contractor B, Contractor A, and for the customer, respectively. Figures 9-4 and 9-5 depict the CDFs and the probability of exceedances (1-CDFs) of the customer and Contractors A and B. Table 9-4 summarizes the variations among the expected value and the conditional expected values (for four different partitionings) generated by the customer and the two contractors. Figure 9-6 depicts these results.

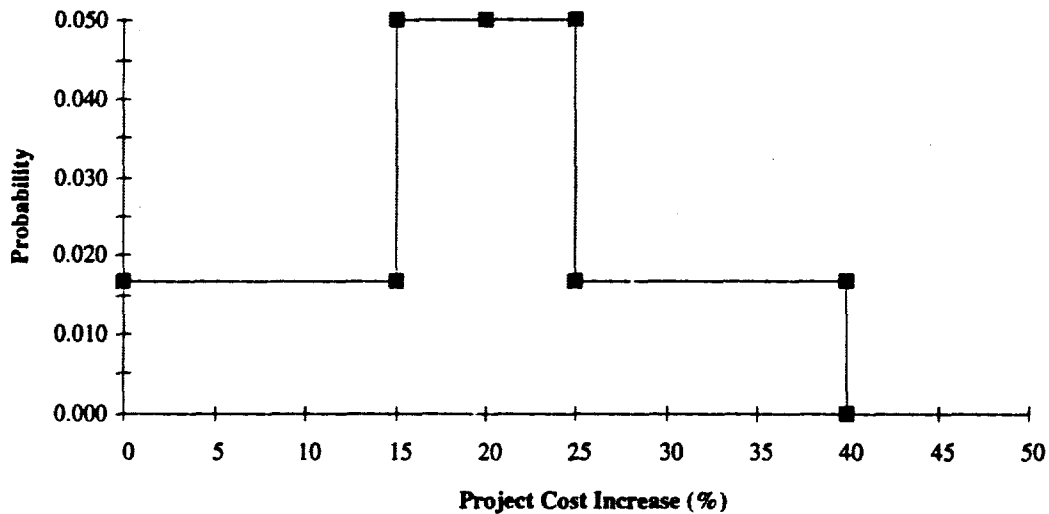


Figure 9-1 PDF of the Projected Cost Increase for Contractor B (Problem 2)

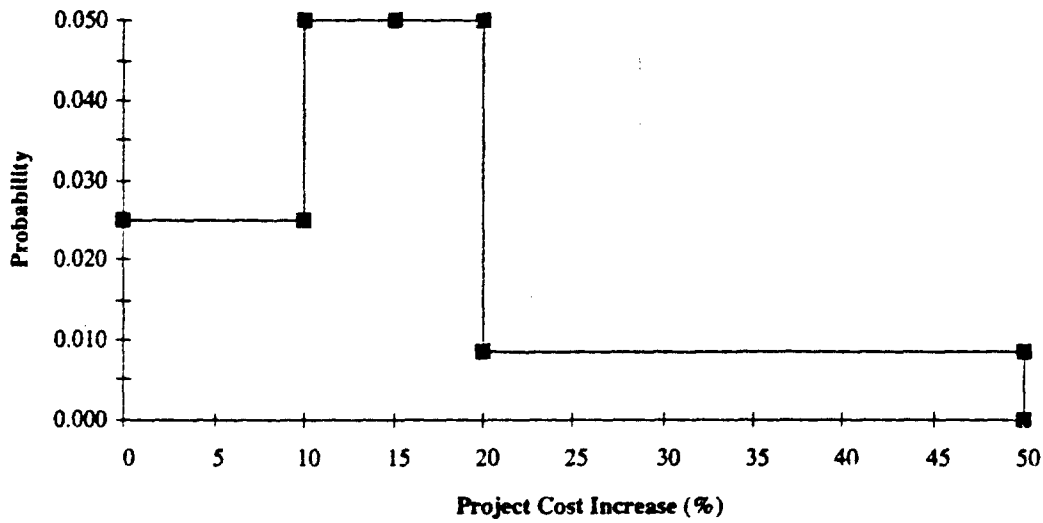


Figure 9-2 PDF of the Projected Cost Increase for Contractor A (Problem 2)

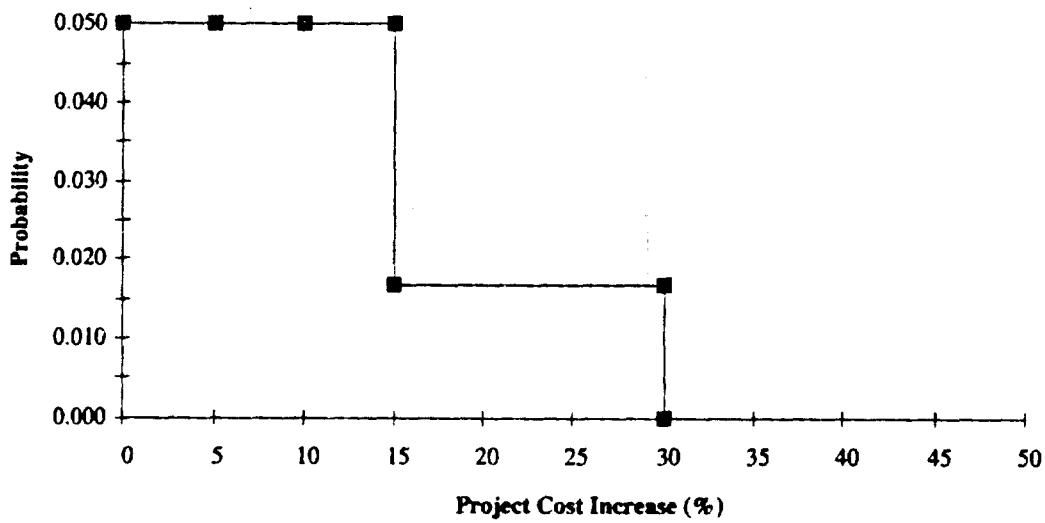


Figure 9-3 PDF for the Customer's Cost Increase (Problem 2)

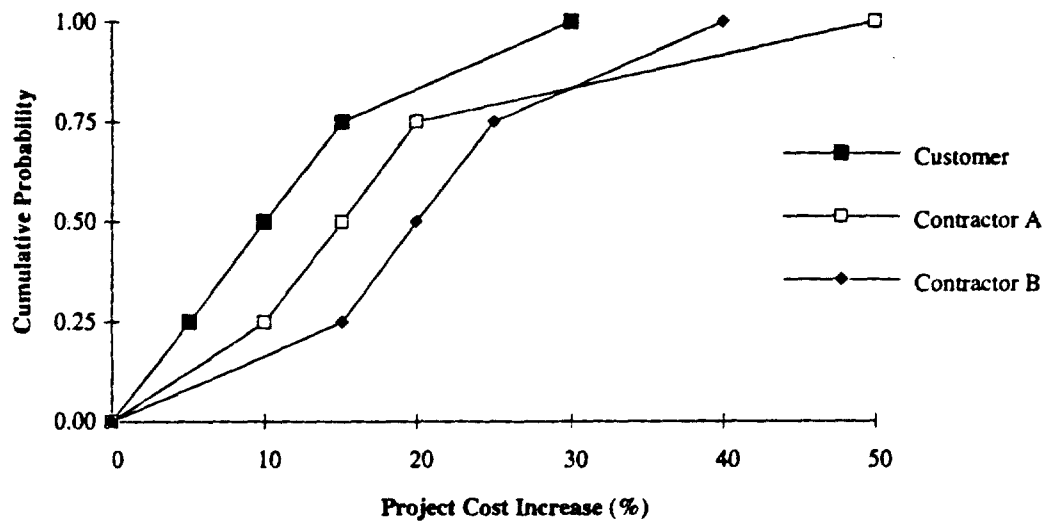


Figure 9-4 CDFs for Project Cost Increases for Contractors A and B and the Customer (Problem 2)

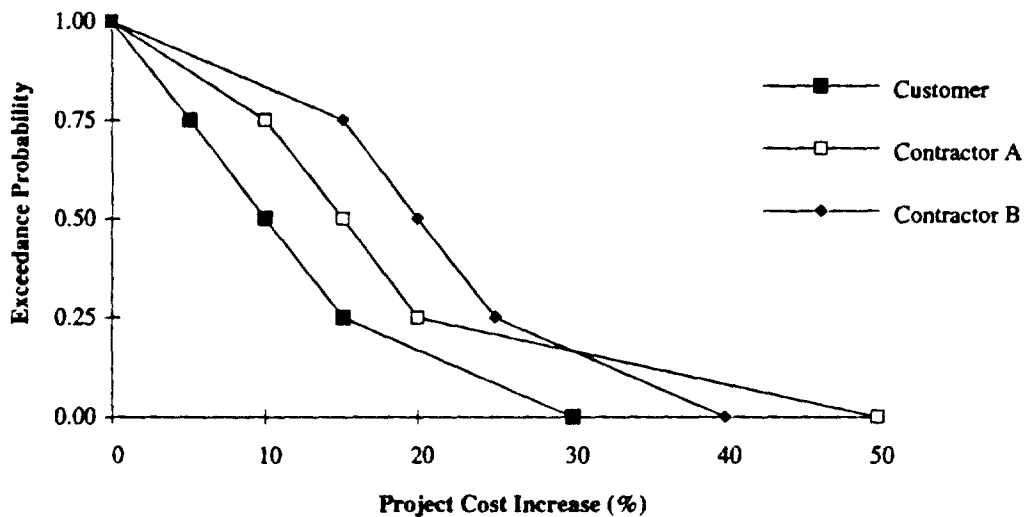


Figure 9-5 Exceedance Probabilities for Project Cost Increases for Contractors A and B and the Customer (Problem 2)

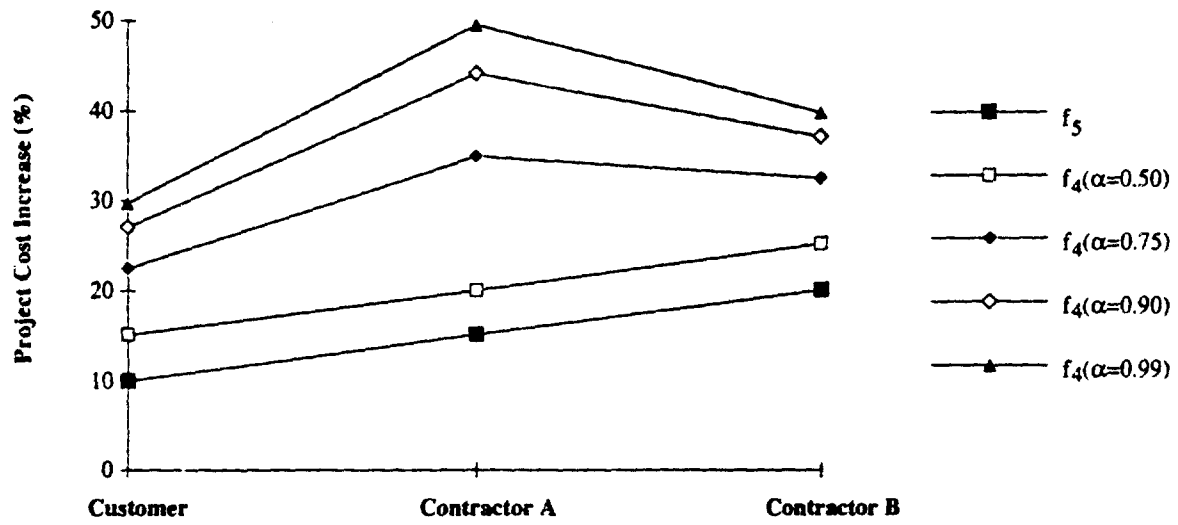


Figure 9-6 Conditional and Unconditional Expected Value of Project Cost Increases for Contractors A and B and the Customer (Problem 2)

Table 9-1 Comparative Tabular CDF (Problem 2)

Fractile	Project Cost Increase (%)		
	Customer	Contractor A	Contractor B
0.00	0	0	0
0.25	5	10	15
0.50	10	15	20
0.75	15	20	25
1.00	30	50	40

Calculations of $f_5(\bullet)$ using the fractile method (Table 9-2):

$$f_5(0,c) = (2.5 + 7.5 + 12.5 + 22.5) 0.25 = 11.25\%$$

$$f_5(1,c) = (5 + 12.5 + 17.5 + 35) 0.25 = 17.50\%$$

$$f_5(2,c) = (7.5 + 17.5 + 22.5 + 32.5) 0.25 = 20\%$$

Calculations of $f_4(0.50; i,c)$ using the fractile method (Table 9-2):

$$f_4(0.50; 0,c) = \frac{0.25 (12.5) + 0.25 (22.5)}{0.5} = 17.5\%$$

$$f_4(0.50; 1,c) = \frac{0.25 (17.5) + 0.25 (35)}{0.5} = 26.25\%$$

$$f_4(0.50; 2,c) = \frac{0.25 (22.5) + 0.25 (32.5)}{0.5} = 27.5\%$$

TABLE 9-2 Summary of Results (Problem 2)

Scenario	Customer i = 0	Contractor A i = 1	Contractor B i = 2
Unconditional Expected value	$f_5(0,c) = 11.25\%$	$f_5(1,c) = 17.50\%$	$f_5(2,c) = 20\%$
1 Partitioning point Corresponding percent of cost increase Conditional expected value	$\alpha = 0.50$ $x = 10.0\%$ $f_4(0.50,0,c)=17.5\%$	$\alpha = 0.50$ $x = 15.0\%$ $f_4(0.50,1,c)=26.25\%$	$\alpha = 0.50$ $x = 20.0\%$ $f_4(0.50,2,c)=27.5\%$
2 Partitioning point Corresponding percent of cost increase Conditional expected value	$\alpha = 0.75$ $x = 15\%$ $f_4(0.75,0,c)=22.5\%$	$\alpha = 0.75$ $x = 20\%$ $f_4(0.75,1,c)=35.0\%$	$\alpha = 0.75$ $x = 25\%$ $f_4(0.75,2,c)=32.5\%$
3 Partitioning point Corresponding percent of cost increase Conditional expected value	$\alpha = 0.90$ $x = 24\%$ $f_4(0.90,0,c)=27.0\%$	$\alpha = 0.90$ $x = 38\%$ $f_4(0.90,1,c)=44.0\%$	$\alpha = 0.90$ $x = 34\%$ $f_4(0.90,2,c)=37\%$
4 Partitioning point Corresponding percent of cost increase Conditional expected value	$\alpha = 0.99$ $x = 29.4\%$ $f_4(0.99,0,c)=29.7\%$	$\alpha = 0.99$ $x = 48.8\%$ $f_4(0.99,1,c)=49.4\%$	$\alpha = 0.99$ $x = 39.4\%$ $f_4(0.99,2,c)=39.7\%$

It is constructive to clarify the information summarized in Table 9-2. Consider the customer's column. According to the customer's estimates (as shown in Figure 9-3), the common, unconditional expected value of cost overrun is 11.25%. Consider Problem 3. From Figure 9-5 or through mathematical calculations on the basis of the information provided by the customer (as shown in Table 9-1), it can be determined that there is 0.1 probability of project cost overrun that would exceed 24% of its original scheduled cost. Thus, the conditional expected value of extreme cost overrun beyond the 24% (or extreme cost overrun with exceedance probability that is below 0.1) is 27%. Finally, consider scenario 4 in the customer's column. There is 0.01 probability ($\alpha = 0.99$, thus $1-\alpha = 0.01$) of project cost overrun that would exceed 29.4% of its original scheduled cost. Thus, the conditional expected value of extreme cost overrun beyond

the 29.4% (or extreme cost overrun with an exceedance probability that is below 0.01) is 29.7%.
Similar interpretations applies to the contractors' columns.

10. Problem 3

The fractile method was used in Problems 1 and 2 in estimating the customer's and the two contractors' variances of project cost. In this example the triangular distribution will be used for such an estimate. Note that in constructing a triangular distribution no estimates of probabilities per se are required. Each contractor is asked to provide three values of the projected cost: (a) lower bound (b); upper bound; and (c) most likely. Table 10-1 provides a summary of these estimates for the customer and for Contractors A and B. On the basis of these estimates, the three PDFs are Constructed in Figures 10-1, 10-2, and 10-3 for the customer and for Contractors A and B, respectively.

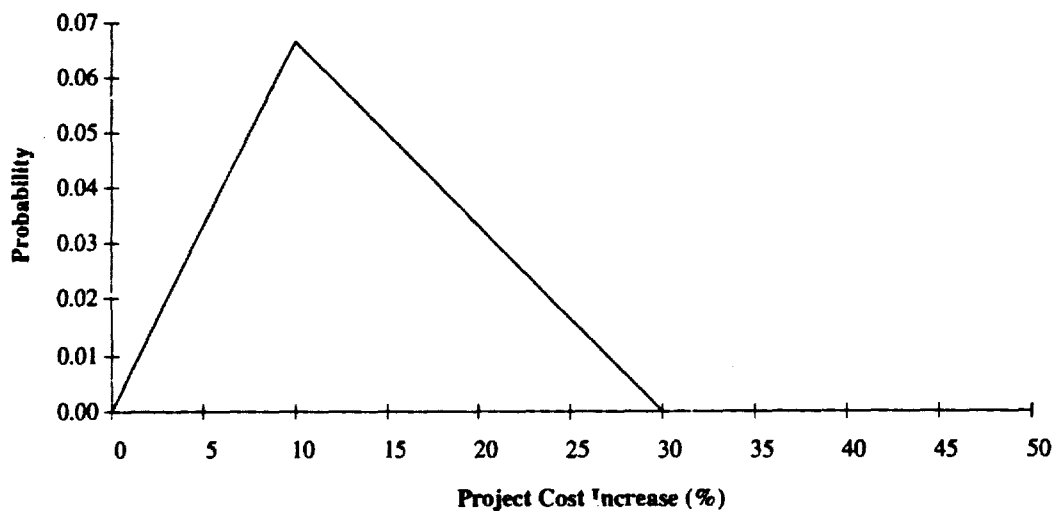


Figure 10-1 PDF for the Customer's Cost Increase (Problem 3)

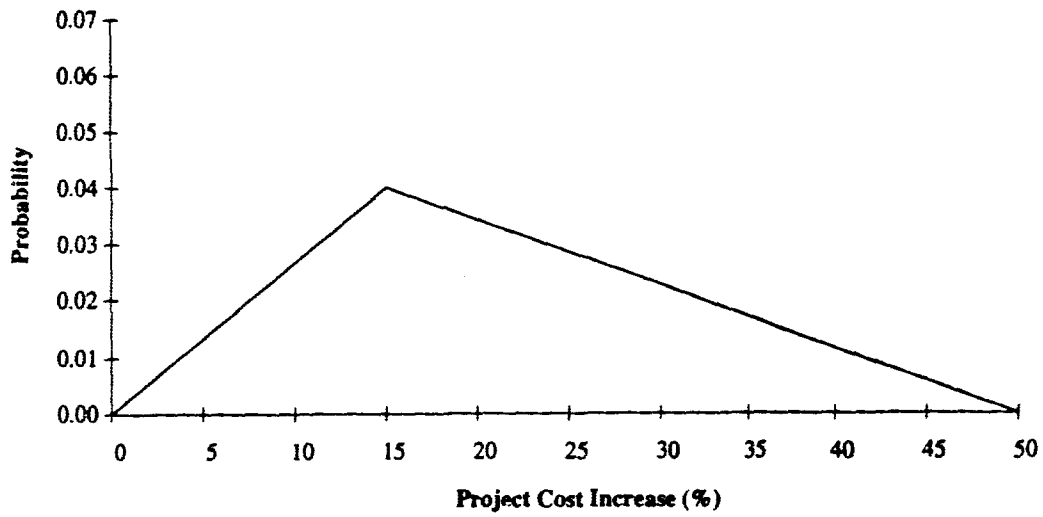


Figure 10-2 PDF for the Project Cost Increase for Contractor A (Problem 3)

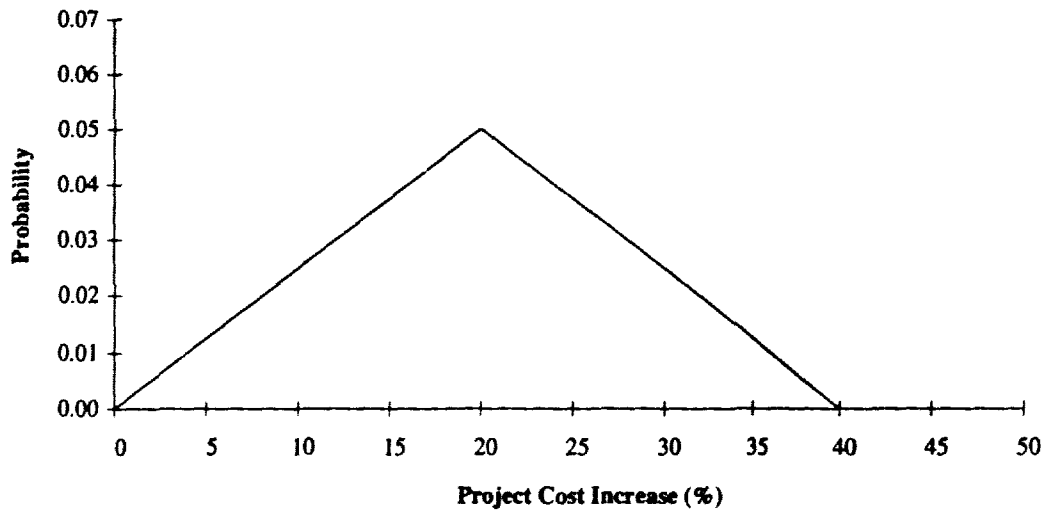


Figure 10-3 PDF for the Project Cost Increase for Contractor B (Problem 3)

Table 10-1 Comparative Assessed Values for the Triangular Distribution (Problem 3)

Assessed Value	Project Cost Increase (%)		
	Customer	Contractor A	Contractor B
Lower Bound	0	0	0
Upper Bound	30	50	40
Most Likely	10	15	20

For comparison, Figure 10-4 depicts all PDFs on the same graph and Figure 10-5 depicts a generic triangular distribution.

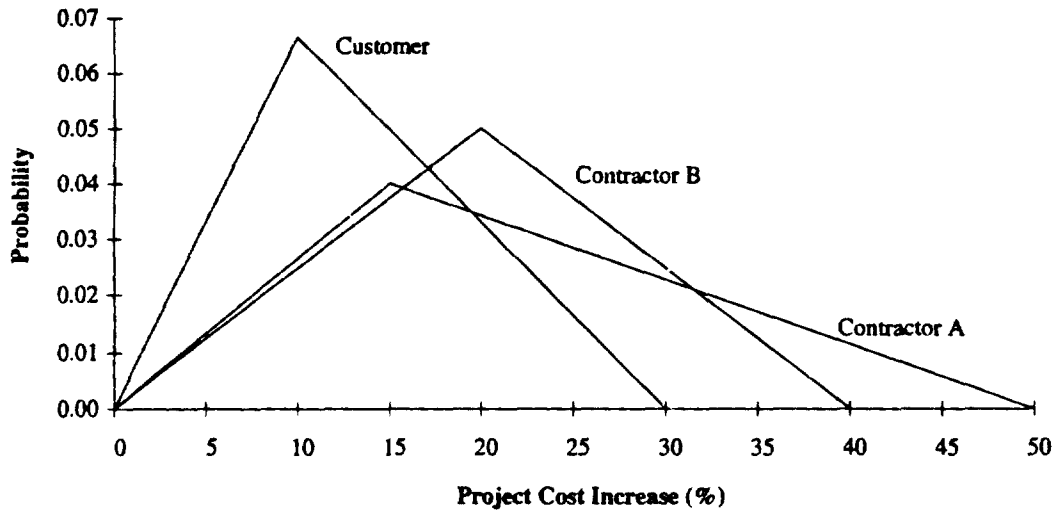


Figure 10-4 Comparative PDFs for Project Cost Increase for Customer and Contractors A and B (Problem 3)

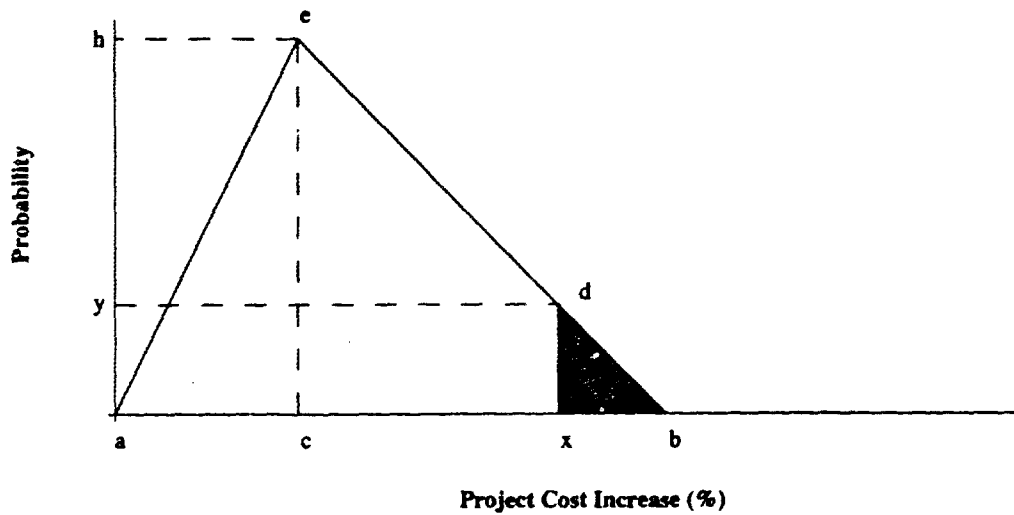


Figure 10-5 Example Triangular Distribution

Table 10-2 summarizes the calculations of the conditional and unconditional expected values for various scenarios for the triangular distribution. Figure 10-6 depicts graphically the results summarized in Table 10-2. The detailed calculations can be found in Appendix D.

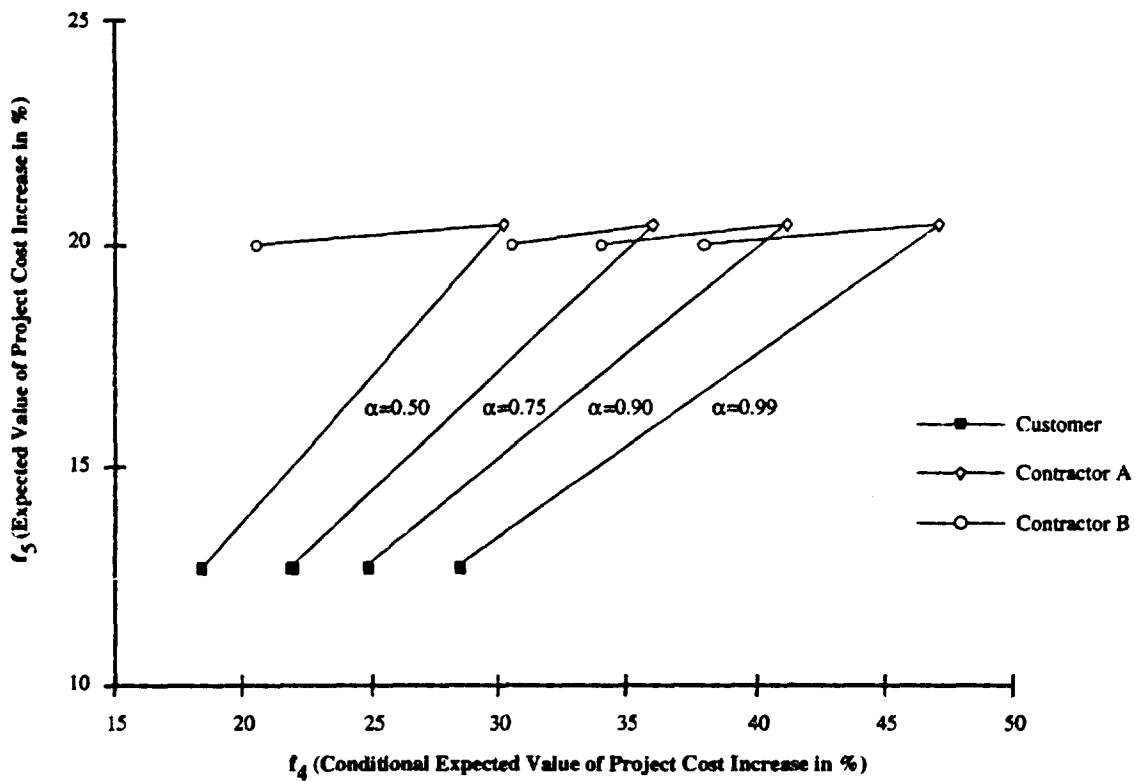


Figure 10-6 Graphical Representation of the Results Summarized in Table 10-2 (Problem 3)

Table 10-2 Summary of Results for Triangular Distribution (Problem 3)

	f_5	$\alpha = 0.50$		$\alpha = 0.75$		$\alpha = 0.90$		$\alpha = 0.99$	
		x	f_4	x	f_4	x	f_4	x	f_4
Percentage Cost Increase									
Customer	13.33	12.68	18.45	17.75	21.84	22.25	24.84	27.55	28.37
Contractor A	21.67	20.42	30.28	29.08	36.06	36.77	41.18	45.82	47.21
Contractor B	20.00	20.00	20.67	25.86	30.57	31.06	34.04	37.17	38.11

11. Comparative Analysis Among Contractors

An important activity within Phase III of the four-phase acquisition process is understanding the reasons and the genesis for the variations of the estimates among the contractors and explaining these differences on the basis of the evidence collected through the Taxonomy, the interviews and other ways. Figure 11-1 represents four possible ways for such variations to occur:

1. The contractor knows what is to be known about the project.
2. The contractor does not know what is to be known about the project.
3. The contractor knows and is aware of the unknowns and uncertainties about the project.
4. The contractor does not know and is not aware of the uncertainties surrounding the project.

Of course this knowledge or the lack thereof is not absolute and the contractor's own knowledge may be at various levels between complete awareness and complete ignorance. This discussion assumes that no gaming is taking place. Safeguards must be developed, however, to secure against a contractor who opts to game the system.

At this stage of the analysis, the Air Force and its team have at their disposal a wealth of information that was either generated by their own team ($i=0$), or by the bidding contractors ($i=1, 2, 3, \dots, n$). Table 8-1 summarizes the symbols and their meanings.

The following comparative analyses are critically important to the contractor's selection process and thus in reducing software nontechnical risk.

$f_5(0, c)$ vs. $f_5(i, c)$ ($i = 1, 2, 3, \dots, n$)

$f_4(\alpha; 0, c)$ vs. $f_4(\alpha; i, c)$ ($i = 1, 2, 3, \dots, n$)

$f_5(0, t)$ vs. $f_5(i, t)$ ($i = 1, 2, 3, \dots, n$)

$f_4(\alpha; 0, t)$ vs. $f_4(\alpha; i, t)$ ($i = 1, 2, 3, \dots, n$)

With these and other comparisons that can be made as desired, the customer's ability to assess the various risks and thus to mitigate and manage them is greatly enhanced. For example, when there appears to be a substantial discrepancy either between the customer's and one of the contractor's estimates or among the estimates of the various contractors, the customer can inquire

(if legally permissible) about evidence and sources of these variations; otherwise, the customer can draw conclusions on the contractor's estimation capabilities and honesty. The use of the SEI Software Risk Taxonomy-Based Questionnaire [SEI 1993] in conjunction with these analyses will be discussed in a subsequent section.

The methodology advocated in this paper does not embrace an adversarial relationship between the customer and the prospective contractors. Project cost overrun and schedule time delay are not assumed to happen necessarily because of contractors' conspiracy or malice. Rather, the premise here is that often the customer and the contractors do not adhere to a systemic risk assessment approach in their evaluation and projection of software nontechnical risk and the unintended result is a cost overrun of delay in project completion schedule.

The use of the Software Risk Taxonomy-Based Questionnaire developed by the Software Engineering Institute constitutes the basis for this needed systemic risk assessment approach [Carr et al. 1993]. The SEI taxonomy questionnaire is divided into three major parts:

- Product Engineering
- Development Environment
- Program Constraints

The taxonomy makes a distinction among known, unknown, and unknowable risks:

The risks in a software development project can be known, unknown, or unknowable. Known risks are those that one or more project personnel are aware of—if not explicitly as risks, at least as concerns. The unknown risks are those that would be surfaced (i.e., become known) if project personnel were given the right opportunity, cues, and information. The unknowable risks are those that, even in principle, no one could foresee. Hence these risks, while potentially critical to project success, are beyond the purview of any risk identification method.

This concept is further expanded in this paper through Figure 11-1, where a distinction among and the impact of knowledge, reality, and perception of uncertainties are tabulated.

	Level of Perceived Risk	Known to the Professional Community	Unknown to the Professional Community
Known to the Contractor	High Risk	<ul style="list-style-type: none"> • Manageable high risk • Know and understand the sources of risk and probability of impact • Realistic assumptions and premises • Risk mitigation strategies are complex 	<ul style="list-style-type: none"> • Manageable high risk • Know and understand the sources of risk • Realistic assumptions and premises • New technology where contractor has competitive advantage • Difficult mitigation strategies (limited outside knowledge to draw from)
	Low Risk	<ul style="list-style-type: none"> • Manageable low risk • Know and understand the sources of risk and probability of impact • Realistic assumptions and premises • Mitigation strategies to circumvent risk are more apparent 	Same as above
Unknown to the Contractor	High Risk	<ul style="list-style-type: none"> • Ill-managed high risk • Unrealistic assumptions and premises • Contractor is not knowledgeable or is not prepared for some phases of the project • Risk mitigation strategies are difficult or unsound • Contractor must rely on outside assistance and training 	<ul style="list-style-type: none"> • Low likelihood of high risk identification and management • Unprecedented technology; high developmental activity • Risk mitigation strategies and understanding of technical uncertainties are difficult • This is a learn-as-you-go type activity
	Low Risk	Same as "Known to the Contractor"	Same as above

Figure 11-1 Knowledge, Reality, and Perception of Uncertainties

The primary focus of the SEI's risk identification process is to elicit known and unknown risks from the personnel associated with the project, e.g., administrative and technical management, development engineers, proposal team, and from cost estimators. The identification process consists of a taxonomy-based questionnaire and a method for conducting interviews using this questionnaire. This enables the interviewers to probe for both technical and nontechnical risks that affect the project. The information that is gathered from the interviews can be grouped and ordered using a set of criteria and a risk ranking and filtering method. The strength of this approach is that the process is repeatable and systematic, and it enables the analysis and comparison of data from multiple organizations. The analysis and comparison of risks and concerns coupled with the extreme events information will provide the customer with a foundation upon which to make more informed decisions regarding the risks in the cost or schedule estimates.

An additional benefit of the analysis is that customers can gain valuable insight into the contractors' assumptions and the depth of their understanding regarding the requirements and technology associated with the project.

Appendix A provides highlights of the Taxonomy.

12. Epilogue

Controlling the cost and time schedule of major projects has been and continues to be a major problem facing government and non-government acquisition managers. Software development projects are no exception because of the close influence and interaction between software technical and nontechnical risks and the diverse sources and causes that constitute the driving force behind these risks, the acquisition manager's job is complicated. One of the major premises of this paper is that a careful, systemic, and analytically-based process for contractor selection is imperative to the prevention of major risks of cost overruns and time delays. This paper proposes such an acquisition process. The four-phase process can be best viewed as a framework rather than as a rigid step-by-step procedure. The obvious limitation in the scope of any single paper prevents a full demonstration of each of the four phases of the proposed acquisition process. The three sample problems should successfully communicate with the reader the mathematical mechanics associated with Phase I and the construction of the measure of risk of extreme events in Phase II. The readers who are more familiar with the SEI Taxonomy-Based Questionnaire will be able to relate more easily with its use in Phases II and III. Similar statements can be made on the familiarity with the risk ranking and filtering method, the independent verification and validation team, and with other methods used in the proposed acquisition process. As a framework, the discussion in this paper must be construed by the reader as the beginning of a dialogue in the direction toward the quantification and management of the risks of cost overruns and time delays associated with software development. In this spirit we consider this paper to be a precursor one which will be followed by others in the future. The expected benefits that result from the prevention of major and extreme risks, combined with the low expected cost of early mitigation strategies, encourage us to believe that this area is worthy of much further consideration.

References

- [Alpert 82] Alpert, M.; & Raiffa, H. "A Progress Report on the Training of Probability Assessors." In *Judgment Under Uncertainty: Heuristics and Biases*. D. Kahneman, P. Slovic, and A. Tversky, eds. Cambridge: Cambridge University Press, 1982.
- [Asbeck 84] Asbeck, E.; & Haimes, Y. Y. "The Partitioned Multiobjective Risk Method." *Large Scale Analysis: An International Journal* 11, 2 (1984).
- [Boehm 81] Boehm, Barry W. *Software Engineering Economics*. Englewood Cliffs, N.J.: Prentice Hall, 1981.
- [Carr 93] Carr, M.J.; Konda, S.L.; Monarch, I.; Ulrich, F.C.; & Walker, C.F. *Taxonomy-Based Risk Identification (CMU/SEI-93-TR-6)*. Pittsburgh, Pa.: Software Engineering Institute, Carnegie Mellon University, 1993.
- [Chittister 93] Chittister, Clyde; & Haimes, Yacov Y. "Risk Associated with Software Development: A Holistic Framework for Assessment and Management." To appear in *IEEE Transactions on Systems, Man, and Cybernetics* 23, 4 (1993).
- [Chittister 94] Chittister, Clyde; & Haimes, Yacov Y. "Assessment and Management of Software Technical Risk." To appear in *IEEE Transactions on Systems, Man, and Cybernetics*, 24, 1 (1994).
- [GAO 92] General Accounting Office. "Embedded Computer Systems: Significant Software Problems on C-17 Must be Addressed" (GAO/IMTEC-92-48). Washington, D.C.: Government Printing Office, May 7, 1992.
- [Haimes 92] Haimes, Y.Y. "Risk of Extreme Events and the Fallacy of the Expected Value" Plenary session paper presented at the IFAC/IFORS/IIASA/TIMS Workshop on Support Systems for Decision and Negotiation Processes, Warsaw, Poland, June 24-26, 1992.
- [Humphrey 87] Humphrey, W.S.; & Sweet, W.L. *A Method for Assessing the Software Engineering Capability of Contractors (CMU/SEI-87-TR-23, ADA 187230)*. Pittsburgh, Pa.: Software Engineering Institute, Carnegie Mellon University, 1987.
- [Kaplan 92] Kaplan, S. "'Expert Information' Versus 'Expert Opinions': Another Approach to the Problem of Eliciting/Combining/Using Expert Knowledge in PRA." *Reliability Engineering and System Safety* 35 (1992): 61-72.
- [Lowrance 76] Lowrance, W.W. *Of Acceptable Risk: Science and the Determination of Safety*. Los Altos, Ca.: William Kaufmann, Inc., 1976.
- [NRC 91] National Research Council. *Innovative Contracting Practices (Transportation Research Circular Number 386)*. Transportation Research Board, December 1991.
- [Smith 88] Smith, Hedrick. *The Power Game: How Washington Works*. New York: Random House, 1988.

APPENDIX A

Taxonomy-Based Risk Identification

(After Carr et al [1993])

Central to the risk identification method is the software development taxonomy. The taxonomy provides a framework for organizing and studying the breadth of software development issues. Hence, it serves as the basis for eliciting and organizing the full breadth of software development risks—both technical and nontechnical. The taxonomy also provides a consistent framework for the development of other risk management methods and activities.

The software taxonomy is organized into three major classes.

1. **Product Engineering.** The technical aspects of the work to be accomplished.
2. **Development Environment.** The methods, procedures, and tools used to produce the product.
3. **Program Constraints.** The contractual, organization, and operational factors within which the software is developed but which are generally outside of the direct control of the local management.

These taxonomic classes are further divided into *elements* and each element is characterized by its *attributes*.

Figure A-1 contains a schematic of the taxonomy. An overview of the taxonomy groups and their hierarchical organization is provided in Figure A-2.

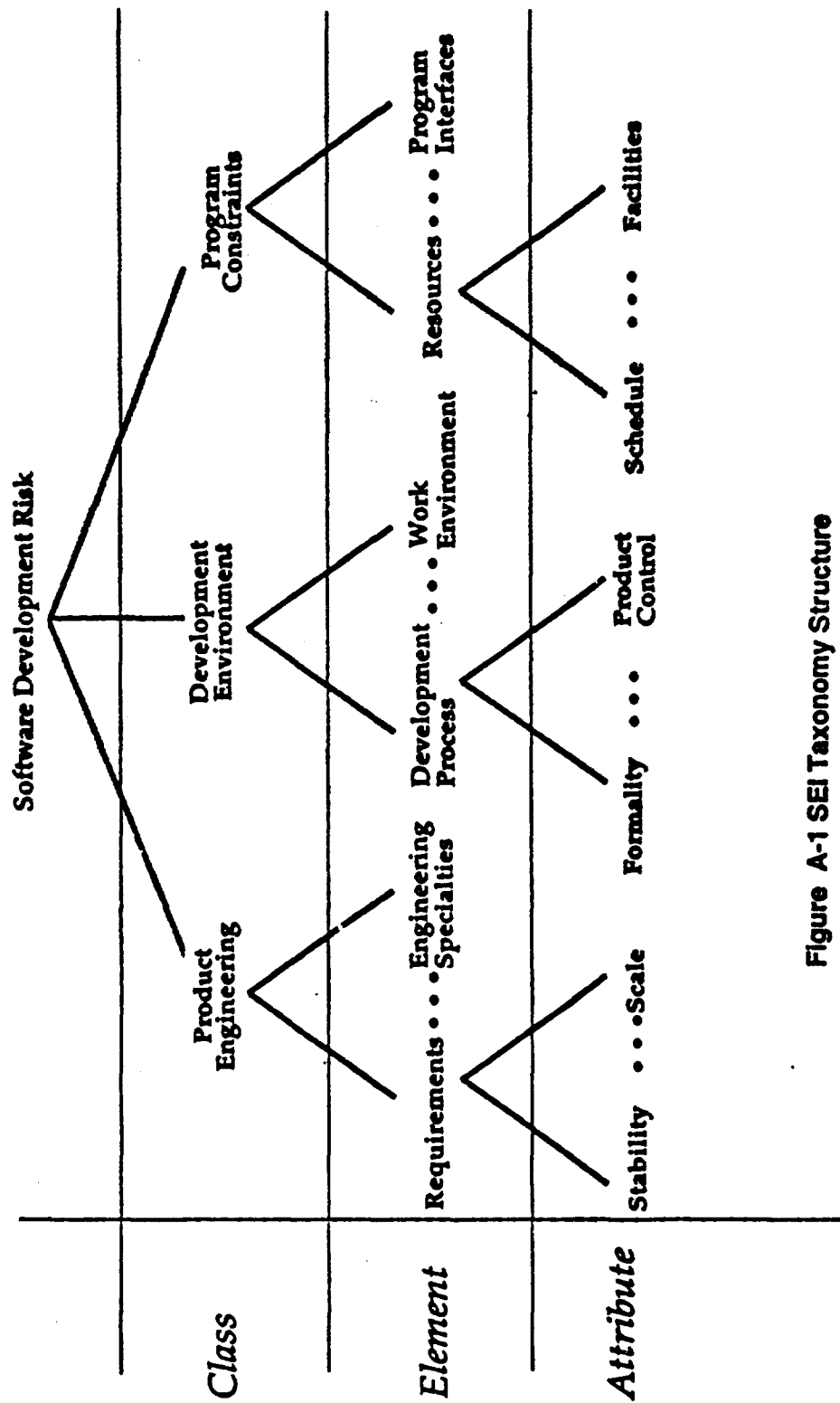


Figure A-1 SEI Taxonomy Structure

A. Product Engineering

1. Requirements
 - a. Stability
 - b. Completeness
 - c. Clarity
 - d. Validity
 - e. Feasibility
 - f. Precedent
 - g. Scale
2. Design
 - a. Functionality
 - b. Difficulty
 - c. Interfaces
 - d. Performance
 - e. Testability
 - f. Hardware Constraints
 - g. Non-Developmental Software
3. Code and Unit Test
 - a. Feasibility
 - b. Testing
 - c. Coding/Implementation
4. Integration and Test
 - a. Environment
 - b. Product
 - c. System
5. Engineering Specialties
 - a. Maintainability
 - b. Reliability
 - c. Safety
 - d. Security
 - e. Human Factors
 - f. Specifications

B. Development Environment

1. Development Process
 - a. Formality
 - b. Suitability
 - c. Process Control
 - d. Familiarity
 - e. Product Control
2. Development System
 - a. Capacity
 - b. Suitability
 - c. Usability
 - d. Familiarity
 - e. Reliability
 - f. System Support
 - g. Deliverability
3. Management Process
 - a. Planning
 - b. Project Organization
 - c. Management Experience
 - d. Program Interfaces
4. Management Methods
 - a. Monitoring
 - b. Personnel Management
 - c. Quality Assurance
 - d. Configuration Management
5. Work Environment
 - a. Quality Attitude
 - b. Cooperation
 - c. Communication
 - d. Morale

C. Program Constraints

1. Resources
 - a. Schedule
 - b. Staff
 - c. Budget
 - d. Facilities
2. Contract
 - a. Type of Contract
 - b. Restrictions
 - c. Dependencies
3. Program Interfaces
 - a. Customer
 - b. Associate Contractors
 - c. Subcontractors
 - d. Prime Contractor
 - e. Corporate Management
 - f. Vendors
 - g. Politics

Figure A-2 - Taxonomy of Software Development Risks

APPENDIX B

Calculating the Expected Value Using Equation 3

The expected value of the percentage of project cost increase using Equation 3 is:

$$\begin{aligned} f_5(\bullet) &= \int_0^{10} xp_X(x) dx + \int_{10}^{20} xp_X(x) dx + \int_{20}^{50} xp_X(x) dx \\ f_5(\bullet) &= \int_0^{10} 0.025x dx + \int_{10}^{20} 0.05x dx + \int_{20}^{50} 0.00833x dx \\ &= 0.025 \frac{x^2}{2} \Big|_0^{10} + 0.05 \frac{x^2}{2} \Big|_{10}^{20} + 0.00833 \frac{x^2}{2} \Big|_{20}^{50} \\ &= 0.025 (50) + 0.05 (200 - 50) + 0.00833 (1250 - 200) \\ &= 1.25 + 7.50 + 8.75 \\ &= 17.50\% \text{ (i.e., total cost of \$ (150 + 26.25) million)} \end{aligned}$$

APPENDIX C

Calculating the Conditional Expected Value Using Equation 5

$$f_4(\bullet) = \frac{\int_{38}^{50} x p(x) dx}{\int_{38}^{50} p(x) dx} = \frac{\int_{38}^{50} x k dx}{\int_{38}^{50} k dx} = \frac{\frac{x^2}{2} \Big|_{38}^{50}}{x \Big|_{38}^{50}}$$

$$f_4(\bullet) = 44\% \text{ (i.e., } \$ (150 + 66) = \$216 \text{ million)}$$

APPENDIX D

Calculating the Conditional and Unconditional Expected Values of Cost Overruns for Problem 3

For the triangles in Figures 10-1, 10-2, and 10-3 to qualify as PDFs, the area under each triangle must be equal to 1. Consider the generic triangular distribution depicted in Figure 10-5. Denote the PDF at any point x by $p(x)$; thus, the PDF at the most likely cost estimate is $p(c)$.

The following functional relationships hold [Law & Kelton 1992]:

$$\text{PDF: } p(x) = \begin{cases} \frac{2(x-a)}{(b-a)(c-a)} & \text{if } a \leq x \leq c \\ \frac{2(b-x)}{(b-a)(b-c)} & \text{if } c < x \leq b \\ 0 & \text{otherwise} \end{cases}$$
$$\text{CDF: } P(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{(x-a)^2}{(b-a)(c-a)} & \text{if } a \leq x \leq c \\ 1 - \frac{(b-x)^2}{(b-a)(b-c)} & \text{if } c < x \leq b \\ 0 & \text{otherwise} \end{cases}$$

The mean (expected value) $f_5(i,c) = \frac{a+b+c}{3}$

The variance = $\frac{a^2 + b^2 + c^2 - ac - ab - bc}{18}$

From geometry:

Area of triangle = $\frac{(b-a)}{2} p(c) = 1$, where a , b , and c are real numbers, and $a < c < b$

Thus, $p(c) = \frac{2}{b-a}$

Thus, for the customer, the probability of c (the most likely cost increase of 10%) is:

$$p(c) = p(10) = \frac{2}{30-0} = 0.067$$

In other words, according to the customer's estimate, the probability of a 10% increase in the cost is 0.067.

$$f_5(0,c) = \frac{0 + 30 + 10}{3} = 13.33\%$$

Thus, according to the customer's estimate, the [unconditional] expected value of cost increase is 13.33%.

$$\text{Area of extreme region} = (1-\alpha) = \frac{b-x}{2} y \quad \text{given } x > c \text{ or } \alpha > 1 - \frac{(b-c)h}{2}$$

where $(1-\alpha)$ is the probability of exceedance.

$$\Rightarrow y = \frac{2(1-\alpha)}{b-x}$$

Taking ratios, we have

$$\frac{h}{y} = \frac{be}{bd} = \frac{b-c}{b-x}$$

$$\Rightarrow y = \frac{h(b-x)}{b-c} \quad \text{given } x > c$$

Equating, we can compute the value of x (partition point on the damage axis) as

$$x = b - \sqrt{\frac{2(1-\alpha)(b-c)}{h}} \quad \text{given } x > c, \alpha > 1 - \frac{(b-c)h}{2}$$

Now we can calculate the percentage of project cost increase that corresponds to a probability of exceedance of 0.1 (1 - 0.9).

$$\begin{aligned} x &= 30 - \sqrt{\frac{2(1-0.9)(30-10)}{0.0667}} \\ &= 22.25\% \text{ for customer} \end{aligned}$$

This result means that the percentage in customer's projected project cost increase that corresponds to a probability of exceedance of 0.1 is 22.25.

The centroid for a right-angled triangle lies $\frac{2}{3}$ along the base from the vertex. Since the extreme region forms a right-angled triangle, we can compute the value of $f_4(0.9,0,c)$. That is, the conditional expected value of percentage cost increase, given an exceedance probability of 0.1. This can also be stated differently: $f_4(0.9,0,c)$ is the conditional expected value of percentage cost increase, given the cost increase will be equal or above 22.25%.

$$f_4(\bullet) = b - \frac{2}{3}(b - x)$$

$$= b - \frac{2}{3} \sqrt{\frac{2(1-\alpha)(b-c)}{h}}$$

$$\text{given } x > c, \text{ or } \alpha = 1 - \frac{(b-c)h}{2}$$

$$f_4(\bullet) = 30 - \frac{2}{3}(30 - 22.25)$$

$$= 24.84\% \text{ (for customer)}$$

Alternatively, we can compute the value of $f_4(\bullet)$ using equation 5

$$f_4(\bullet) = \frac{\int_x^{\infty} x p(x) dx}{\int_x^{\infty} p(x) dx}$$

$p(x)$ is given by the slope of the line. From Figure 10-5, we have

$$p(x) = \frac{h}{b-c} (b-x)$$

$$f_4(\bullet) = \frac{\int_x^b x \frac{h(b-x)}{b-c} dx}{\int_x^b \frac{h(b-x)}{b-c} dx}$$

$$= \frac{\int_x^b bx dx - \int_x^b x^2 dx}{\int_x^b b dx - \int_x^b x dx}$$

$$= \frac{b \frac{x^2}{2} \Big|_x^b - \frac{x^3}{3} \Big|_x^b}{bx \Big|_x^b - \frac{x^2}{2} \Big|_x^b}$$

$$= \frac{\frac{b^3 - bx^2}{2} - \frac{b^3 - x^3}{3}}{b^2 - bx - \frac{b^2 - x^2}{2}}$$

$$= \frac{3b^3 - 3bx^2 - 2b^3 + 2x^3}{6b^2 - 6bx - 3b^2 + 3x^2}$$

$$= \frac{b^3 - 3bx^2 + 2x^3}{3b^2 - 6bx + 3x^2}$$

$$\begin{aligned}
&= \frac{b^3 + 2x^3 - 3bx^2}{3(b-x)^2} \\
&= \frac{(30)^3 + 2(22.25)^3 - 3(30)(22.25)^2}{3(30 - 22.25)^2} \\
&= \frac{(27,000 + 22,030.28 - 44,555.625)}{180.1875} \\
&= \frac{(4,474.655)}{180.1875} = 24.83\% \text{ (for customer)}
\end{aligned}$$

Table 10-2 summarizes the results for triangular distribution and Figure 10-6 depicts graphically the results summarized in Table 10-2.

REPORT DOCUMENTATION PAGE				
1a. REPORT SECURITY CLASSIFICATION Unclassified		1b. RESTRICTIVE MARKINGS None		
2a. SECURITY CLASSIFICATION AUTHORITY N/A		3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for Public Release Distribution Unlimited		
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A				
4. PERFORMING ORGANIZATION REPORT NUMBER(S) CMU/SEI-93-TR-28		5. MONITORING ORGANIZATION REPORT NUMBER(S) ESC-TR-93-202		
6a. NAME OF PERFORMING ORGANIZATION Software Engineering Institute	6b. OFFICE SYMBOL (if applicable) SEI	7a. NAME OF MONITORING ORGANIZATION SEI Joint Program Office		
6c. ADDRESS (city, state, and zip code) Carnegie Mellon University Pittsburgh PA 15213		7b. ADDRESS (city, state, and zip code) HQ ESC/ENS Hanscom Air Force Base, MA 01731-2116		
8a. NAME OF FUNDING/SPONSORING ORGANIZATION SEI Joint Program Office	8b. OFFICE SYMBOL (if applicable) ESC/ENS	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F1962890C0003		
8c. ADDRESS (city, state, and zip code) Carnegie Mellon University Pittsburgh PA 15213		10. SOURCE OF FUNDING NOS.		
		PROGRAM ELEMENT NO 63756E	PROJECT NO. N/A	TASK NO. N/A
		WORK UNIT NO. N/A		
11. TITLE (Include Security Classification) An Acquisition Process for the Management of Risks of Cost Overrun and Time Delay Associated with Software Dev.				
12. PERSONAL AUTHOR(S) Yacov Y. Haimes, Clyde Chittister				
13a. TYPE OF REPORT Final	13b. TIME COVERED FROM TO	14. DATE OF REPORT (year, month, day) December 1993		15. PAGE COUNT 67 pp.
16. SUPPLEMENTARY NOTATION				
17. COSATI CODES		18. SUBJECT TERMS (continue on reverse of necessary and identify by block number)		
FIELD	GROUP	SUB. GR.	acquisition process	
			risk	
			cost overrun	
19. ABSTRACT (continue on reverse if necessary and identify by block number)				
<p>The ability to quantify risk is essential to the processes of budgeting and scheduling. During the process of hiring to complete specified tasks, customers must be able to verify contractor estimates and to make sound judgements on the risks of cost overruns and time delays. The following two questions are central to this paper: Do developers with little experience over-estimate or underestimate the complexity of the task because of their past experience, the assumptions they make, the models they select, and how they define the model parameters? What are the sources of risk associated with project cost estimation? How can such risk be quantified? To address these questions, this paper proposes a systematic acquisition process that is aimed at assessing and managing the risks of cost overruns and time delays associated with software development.</p> <p>The proposed acquisition process, which is composed of four phases (listed below), is grounded on the following three</p> <p style="text-align: right;">(please turn over)</p>				
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED <input checked="" type="checkbox"/> SAME AS RPT <input type="checkbox"/> DTIC USERS <input checked="" type="checkbox"/>			21. ABSTRACT SECURITY CLASSIFICATION Unclassified, Unlimited Distribution	
22a. NAME OF RESPONSIBLE INDIVIDUAL Thomas R. Miller, Lt Col, USAF		22b. TELEPHONE NUMBER (include area code) (412) 268-7631	22c. OFFICE SYMBOL ESC/ENS (SEI)	

ABSTRACT — continued from page one, block 19

basic premises: a) any single-value estimate of cost or completion time is inadequate to capture and represent the variability and uncertainty associated with cost and schedule. Probabilistic quantification is advocated, using, in this paper, the fractile method and triangular distribution. b) The common expected value when used as a measure of risk, is inadequate; further, if used as the sole measure of risk, it may lead to inaccurate results. The conditional expected value of risk of extreme events is adopted to supplement and complement the common unconditional expected value. c) Probing the sources of risks and uncertainties associated with cost overruns and time delays in software development is essential for the ultimate management of technical and nontechnical risks. The Taxonomy-Based Questionnaire developed by the Software Engineering Institute is adopted.

These basic premises have led to the development of the following four phases in the proposed acquisition process: Phase I, constructing the probability density functions; Phase II, probing the sources of risks and uncertainties; Phase III, analyzing and comparing the significance and validity of the contractors' assumptions and premises regarding the likelihood of technical and nontechnical risks; and Phase IV, drawing conclusions on the basis of the accumulated evidence and ultimately selecting the contractors most likely to complete the project without major cost overruns or time delays. The three example problems are presented to demonstrate the construction of the probability density functions in Phase I and to explain in a more general way the effort involved in Phases II and IV.