Interconnected Estimating Relationships: Their Derivation and Application

ISPA/SCEA Joint Annual Conference & Training Workshop
Albuquerque NM, 7-10 June 2011

Stephen A. Book, MCR LLC, El Segundo CA
Amanda J. Feather, MCR LLC, El Segundo CA

Abstract

By the term “Interconnected Estimating Relationships,” we mean estimating relationships for hardware and software costs, schedules, weights, and below-the-line programmatic costs that are jointly impacted by each other or by the same drivers. There are two common examples of this phenomenon: (1) cost-estimating relationships (CERs) and schedule-estimating relationships (SERs), which are interconnected for many reasons, but primarily because a project’s schedule is a significant driver of its cost; and (2) CERs for hardware and software and techniques for estimating below-the-line costs, the latter of which is typically done by applying a factor or percentage to the hardware and/or software cost estimate. A critical consideration in all of these estimating relationships is the fact the results of such estimates, due to influences of risk and uncertainty, cannot be expressed as single numbers, but rather must be reported as probability distributions. As R.P. Covert (*Journal of Cost Analysis and Parametrics*, Spring 2008, and ISPA/SCEA Joint International Conference, May 2008) observed, not only are cost and schedule estimates uncertain, but the same is true for cost and schedule drivers, such as weight, power, and lines of code. It follows that the drivers are themselves better expressed as probability distributions. A situation that more clearly illustrates the fact that a cost or schedule driver should really be modeled as a probability distribution is the case of so-called “cost-on-cost” CERs, where a below-the-line cost, e.g., system engineering, is estimated as a factor times the cost estimate (not times the cost, because we don’t know the cost – all we know is the estimate). A cost-on-cost CER is really a "cost-estimate-on-cost-estimate" CER, because the driver, namely hardware and/or software cost, is itself the result of an estimating relationship. That estimating relationship drives the system-engineering cost through the estimating relationship $Y = aX$, where Y is system-engineering cost and X is hardware and/or software cost. This presentation offers the statistical foundations of working with interconnected estimating relationships and provides an example of how they operate in practice to give us a better understanding of project costs and schedules.

Acknowledgment

The authors would like to thank William E. Dorman of the AFCAA team located at USAF SMC for interesting and useful discussions on the subject.

We would also thank the members of MCR's internal quality review team – Dr. Neal D. Hulkower, Ms. Samantha J. Cheng, and Messrs. Raymond P. Covert and Aziz N. Sirkhan – not only for their valuable comments and suggestions for improvement, but also their careful reading of the text, which caught several errors of various kinds we would have been embarrassed to present in a public briefing.

Contents

- Current Status of Estimating Relationships
- Why Cost and Schedule Must be Modeled as Probability Distributions
- Why Cost and Schedule Drivers Must be Modeled as Probability Distributions
- What's Wrong with the Way Cost-on-Cost Estimating Relationships are Modeled and How to Fix It
- Applications
- Summary

Contents

Estimating Relationships Today

- **Estimating Relationships are (usually) Algebraic Formulas that Express Items to be Estimated as Functions of "Drivers"**
  – **Cost-Estimating Relationships (CERs)**
  – **Schedule-Estimating Relationships (SERs)**
- **Among Drivers of Cost and Schedule are Included …**
  – **Technical Drivers**
    - **Component, Subsystem, Assembly Weights**
    - **Number of Software Lines of Code**
    - **Data-Processing Requirements**
    - **Power Requirements**
    - **Antenna Diameter**
    - **Signal Strength**
  – **Programmatic Drivers**
    - **Technical Readiness Level (TRL)**
    - **Design Uncertainties**
    - **Probable Schedule Slips of Unknown Duration**
    - **Test Failures and Other Unforeseen Events**

Examples of Today's Estimating Relationships

- **Examples of Symbols**
  - *C*: Cost Estimate
  - *S*: Schedule Estimate
  - *W*: Weight Specification
  - *F*: Frequency Specification
  - *SE*: Systems Engineering Cost Estimate
  - *SLOC*: Source Lines of Code Estimate
- **Examples of Estimating Relationships**
  - $C = 47.34 + 0.67W$ (linear relationship)
  - $C = 5.14W^{0.8}F^{1.4}$ (power relationship)
  - $SE = 0.67C$ (factor relationship)
  - $C = 87.19 + 6.28W^{0.92}$ (triad relationship)
  - $S = 13.6 + 2.1SLOC^{1.27}$ (triad relationship)

Contents

- Current Status of Estimating Relationships
- Why Cost and Schedule Must be Modeled as Probability Distributions
- Why Cost and Schedule Drivers Must be Modeled as Probability Distributions
- What's Wrong with the Way Cost-on-Cost Estimating Relationships are Modeled and How to Fix It
- Applications
- Summary

Fact: Costs* Have Probability Distributions

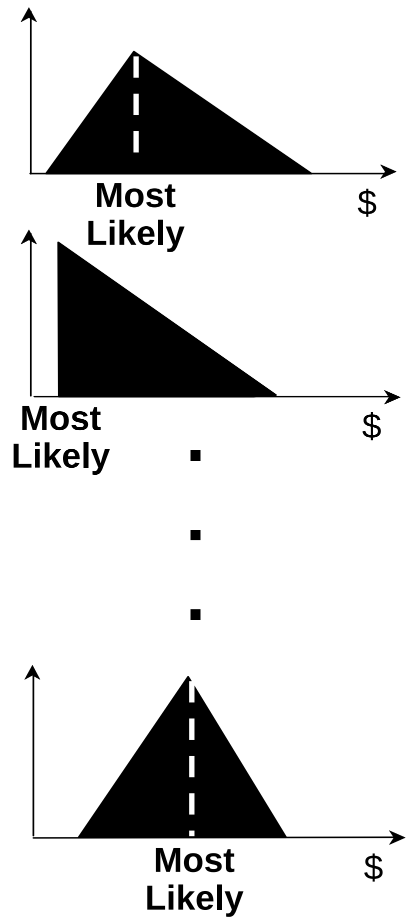
- "Actual" Project Cost is an Uncertain Quantity (Technically, a "Random Variable")
- The "Point" Estimate is not the Only Possible Estimate – There are Others
- The "Best" Estimate is not the Only Possible Estimate Either – Other Estimates are Presumably "Worse"
- Common Use of Phrase "Most Likely" or "Most Probable" Cost Implicitly Assumes that Other Cost Levels are "Less Likely" or "Less Probable"
- This Whole Discussion Implies that Costs Really are Probabilistic in Nature

* . . . and Schedules

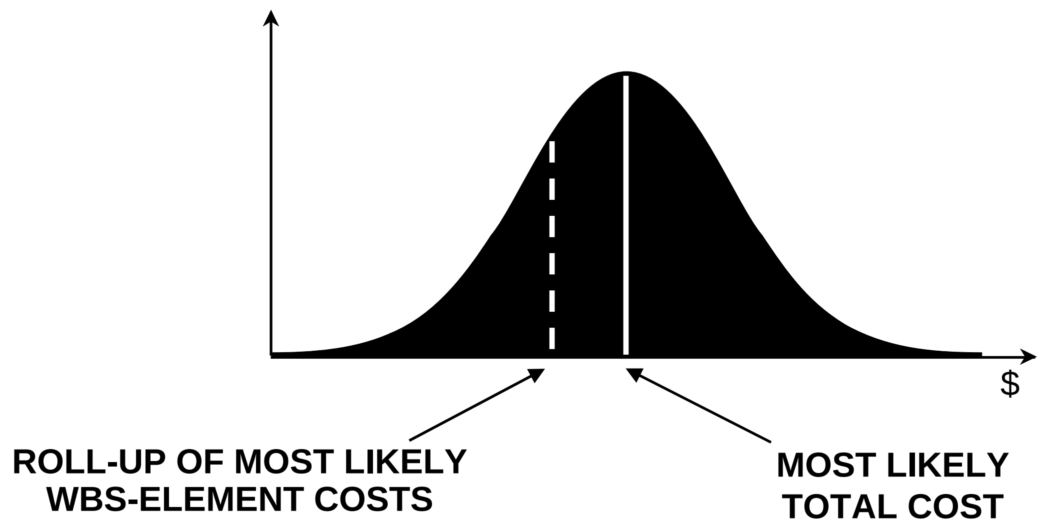
… and “Most-Likely” Costs Don’t Roll Up to the Most-Likely Cost

CRITICAL THINKING. SOLUTIONS DELIVERED.

WBS-ELEMENT TRIANGULAR COST DISTRIBUTIONS



MERGE WBS-ELEMENT COST DISTRIBUTIONS INTO TOTAL-COST NORMAL DISTRIBUTION



THE AEROSPACE CORPORATION

So What Do We Need to Do?

- Treat Every Estimating Task as a Risk Analysis
  - Recognize Uncertainty Inherent in “Best” Estimate and in Every Other Estimate of Cost or Schedule
  - Apply Assessment of Process Risks to Form Probability Distributions for Each Element over Activity
- No Roll-Up: Statistically Sum Element or Activity Costs (Via Monte-Carlo or Analytic Approximation)
  - Avoid Meaningless Outcome of Roll-up Procedure
  - Get Mean, Median, Mode of Total Cost or Schedule
  - Get All Percentiles of Total Cost or Schedule
- Be in the Business of Estimating Probabilities, not Costs or Schedules
  - No One Will Be Able to Say: “Your Estimate Must Be Wrong. We Don’t Have that Much Money/Time in the Budget/Plan.”
  - No Matter What Happens to be in the Budget or Plan, You’ll be Able to Assign a Level of Confidence to It

Contents

- Current Status of Estimating Relationships
- Why Cost and Schedule Must be Modeled as Probability Distributions
- Why Cost and Schedule Drivers Must be Modeled as Probability Distributions
- What's Wrong with the Way Cost-on-Cost Estimating Relationships are Modeled and How to Fix It
- Applications
- Summary

Estimating by Factors

- **Dollars-per-Pound (Factor) CERs Have a Venerated History in "Quickie" Cost Estimating**
  - **The CER Has the Form $C = aW$, where C is Expressed in Dollars, W in Pounds, and a (the factor) in "Dollars per Pound"**
  - **During the SBIRS Source Selection in 1996, a Dollars-per-Pound Estimate based on DSP Block I Showed that the Latter's Cost was Triple the Proposed and Eventually Contracted SBIRS Cost**
- **More Commonly, "Below-the-Line" (BTL) (e.g., SE) Costs are Estimated Using a Factor Relationship, such as $SE = 0.23C$**
  - **Here C is a Random Variable (namely, a quantity that has a probability distribution) as We Have Discussed Earlier**
  - **Therefore SE is also a Random Variable with a Probability Distribution that is a Consequence of the Probability Distribution of C**

Fact: SE Cost Has a Probability Distribution

- In the Factor Relationship $SE = 0.23C$, because C is a Random Variable (namely, a quantity that has a probability distribution), then SE is also a Random Variable with a Probability Distribution
  - Suppose C has a Lognormal Distribution with Mean 500 and Standard Deviation 100
  - Then C's Actual Value is One of Many, Many Possible Values
  - Therefore SE's Actual Value Will be One of Many, Many Possible Values, Namely 0.23 Times Whatever the Actual Value of C Turns out to be
  - This Means that SE has a Probability Distribution Related to the Probability Distribution of C
- It Gets Worse: The "Factor" (0.23 in our example) also has a Probability Distribution, of which 0.23 is Only One Possible Value – More about this Later
- So SE's Cost Probability Distribution is Really the Product of Two Probability Distributions

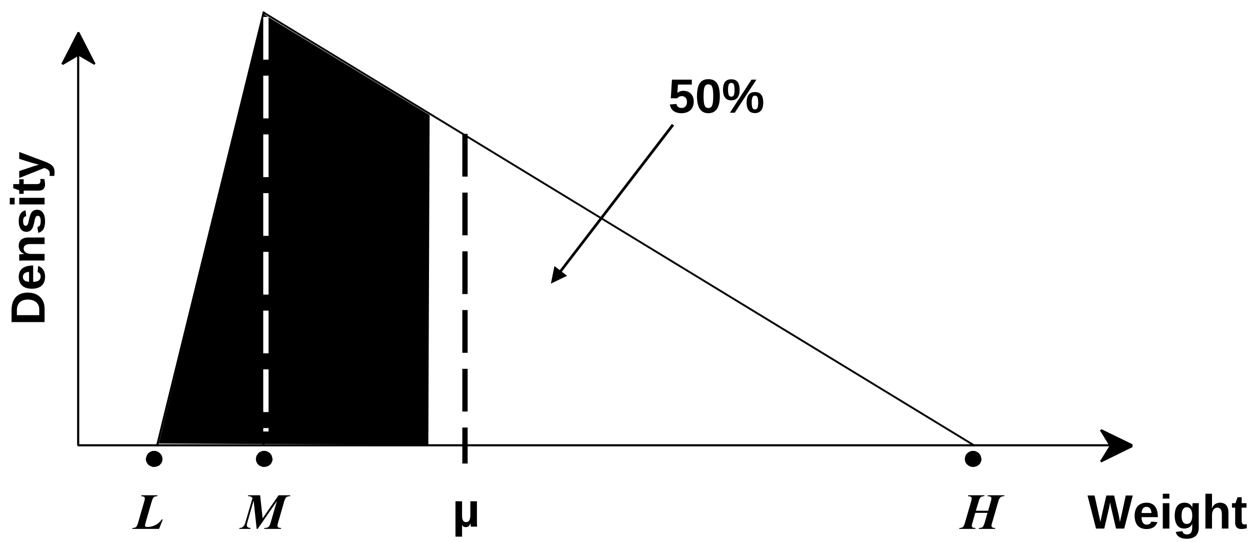
Oh Well, I Guess We Should Come Completely Clean

- **Actual Weight, Frequency, Lines of Code, and Other Cost Drivers are also Random Variables at the Time an Estimate is Made**
  - As a Point of Fact, None of These Cost (and Schedule, for that matter) Drivers is Known Exactly at the Time We are Doing the Estimate
  - They are Only Estimates Themselves of What the System's Developers Believe the Values of Weight, Frequency, Lines of Code, and Other Cost Drivers Will Have to be (the numbers in the CARD*) in Order for the Proposed System to Meet its Performance Specifications
  - Being Estimates, They have Probability Distributions
- **This Fact is Behind the Errors-in-Variables (EIV) Estimating Relationship Idea due to R.P. Covert [References 3 and 4]**

* A Cost Analysis Requirements Description (CARD) is a project-defining document that provides numerical and other estimates of technical and programmatic characteristics of the project.

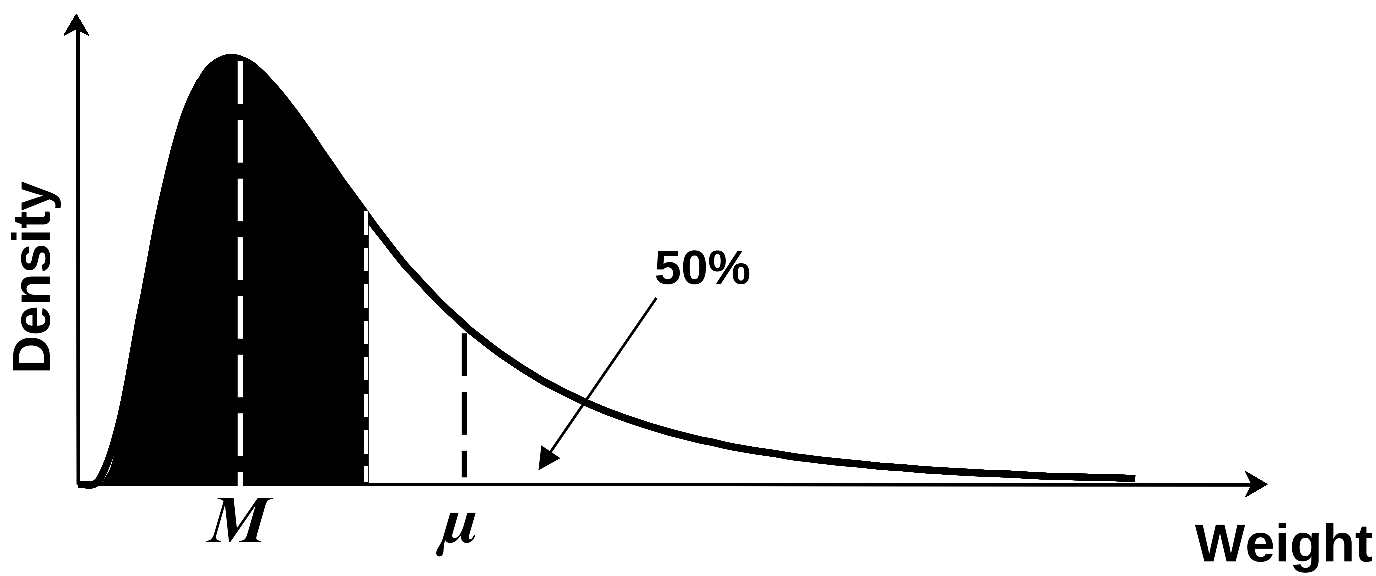
So What Do We Do About This?

- No Problem! We Simply Apply What We Already Know About the Role of Probability Distributions in Cost Estimating
- If Nothing More Specific is Known, We Can Model Weight, for Example, as a Triangular Random Variable Characterized by its L, M, and H Values



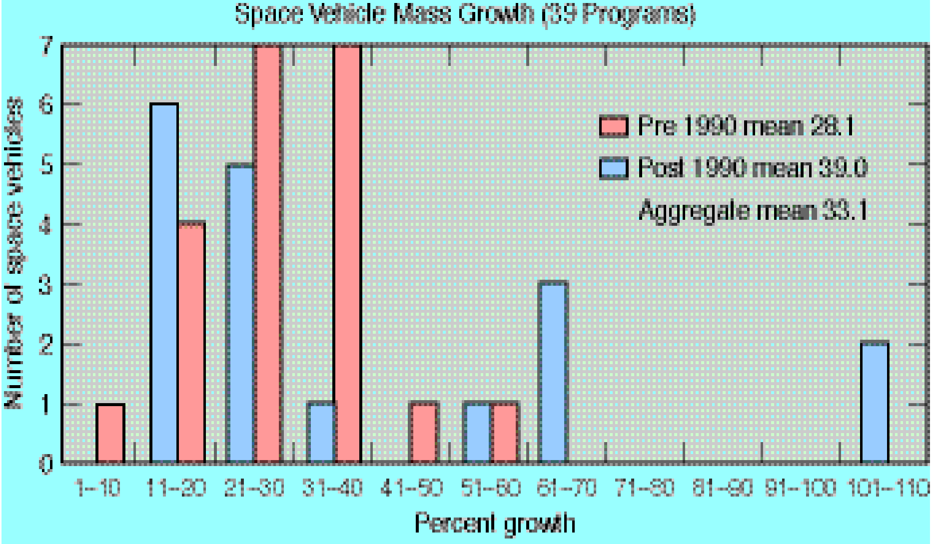
What if We Have More Specific Knowledge?

- **No Problem! Suppose We Have Reason to Believe that the Cost Driver Has a Lognormal Distribution with Mean μ and Standard Deviation σ**
- **Then We Can Model the Driver as a Lognormal Random Variable:**



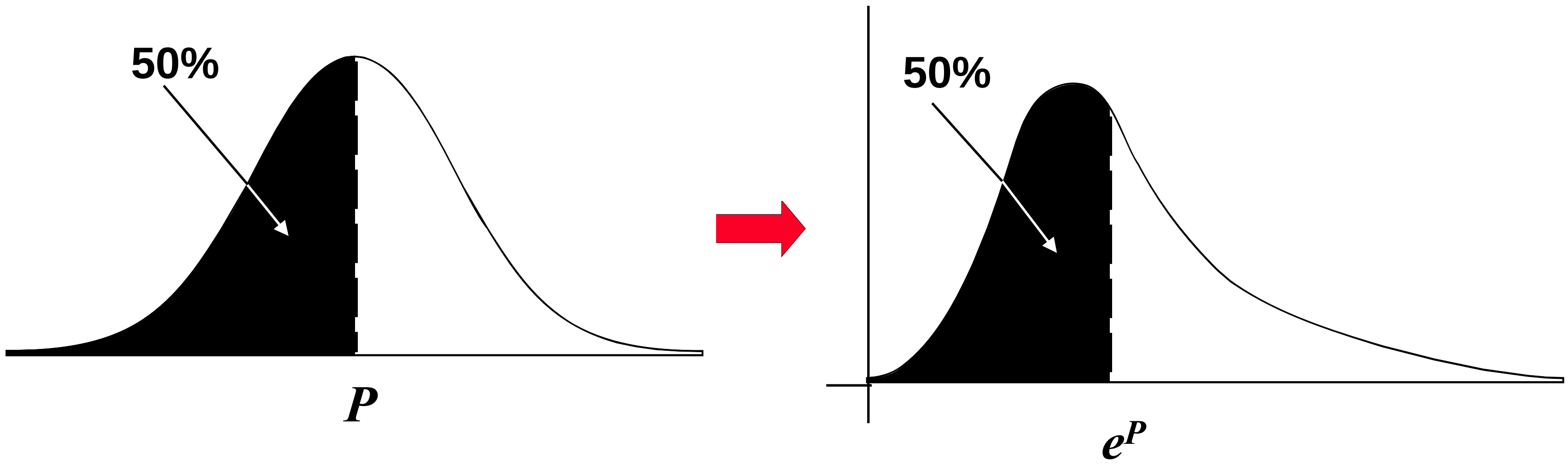
Satellite Weight-Growth Historical Experience

- **A 2006 Aerospace Corp. Report by R. Sugiyama and L. Yang [Reference 11] Found Maximum Satellite Weight Growth Experienced Up to That Time to be 106% at the Extreme Worst for their Data Set**



- **From the Graph, 106% Appears to be Large Even for the Post-1990 Timeframe – This Could Indicate Lognormal, Rather than Triangular Behavior**

Origin of the Lognormal Distribution



X $Y = e^X$

Normal Distribution   ***Lognormal Distribution***

Statistics of Lognormal Distributions

- A Lognormal Distribution with Mean μ and Standard Deviation σ is Derived from a Normal Distribution with Mean P and Standard Deviation Q, so that the Numerical Values of Lognormal Statistics are Derived from the Numerical Values of Normal Statistics
- Lognormal Mode = $M = e^{P - Q^2}$
- Lognormal Mean = μ = $e^{P + \frac{1}{2}Q^2}$
- Lognormal Median = 50th Percentile = $m = e^{P}$
- $(1-\alpha)^{th}$ Percentile = $e^{P + z_\alpha Q}$ (= dollar value at which $P\{Cost \le L_{1-\alpha}\} = 1-\alpha$, where z_α is $(1-\alpha)^{th}$ percentile of the standard normal distribution)
- Lognormal Standard Deviation = σ = $e^{P+\frac{1}{2}Q^2}\sqrt{e^{Q^2} - 1}$

Normal Statistics Expressed in Terms of Lognormal Statistics

- Going Backwards, We Can Consider that a Normal Distribution with Mean P and Standard Deviation Q is Derived from a Lognormal Distribution with Mean μ and Standard Deviation σ, so that the Numerical Values of Normal Statistics are Derived from the Numerical Values of Lognormal Statistics
- For Example:
  - Normal Mean = Normal Median = Normal Mode =

$$P = \frac{1}{2} \lambda n \left\{ \frac{\mu^4}{\left(\mu^2 + \sigma^2\right)} \right\}$$

  - Normal Standard Deviation = $Q = \left\{ \lambda n \left(1 + \frac{\sigma^2}{\mu^2} \right) \right\}^{\frac{1}{2}}$

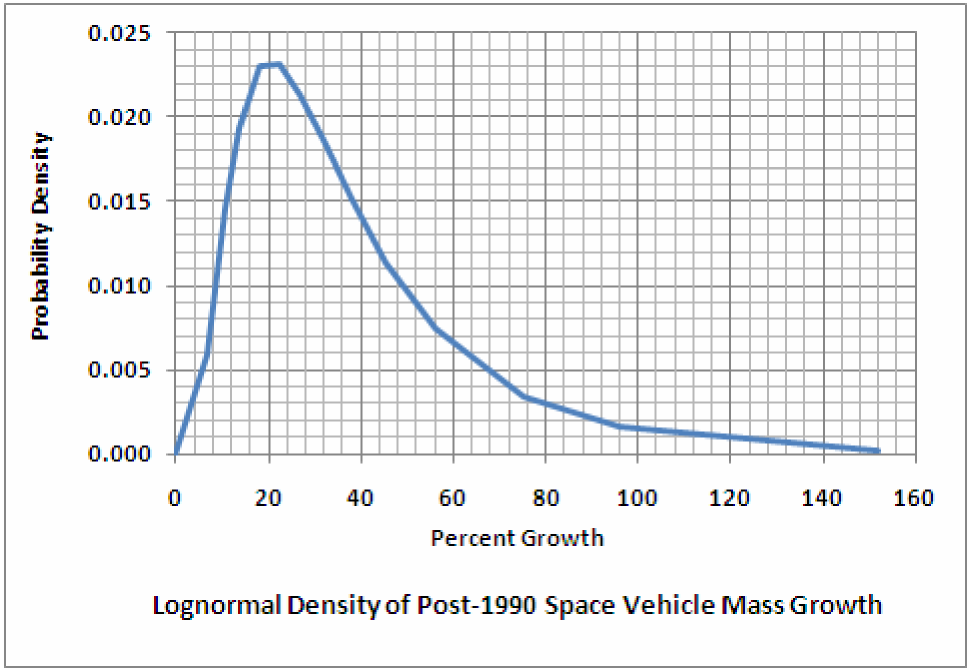
Approximating the Probability Distribution of "Weight"

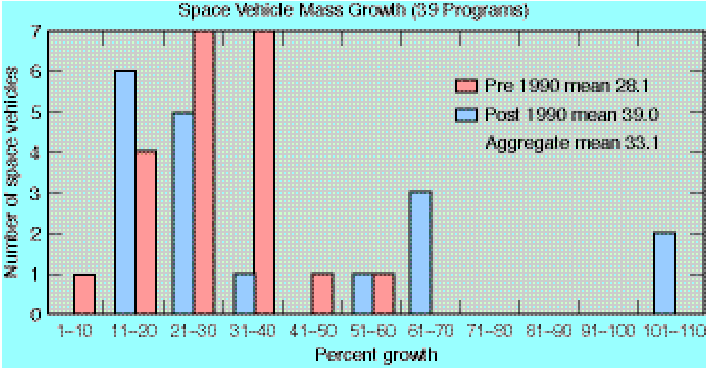
- We Don't Have Access to the Sugiyama-Yang Data Set, but only to their Bar Graph (histogram)
- Therefore we Can only Approximate Distributional Characteristics of Post-1990 Weight ($\approx$ Mass) Growth
- The Table of Calculations on the Next Chart Allow Us to Approximate the Mean and Standard Deviation of the Post-1990 Weight Growth
- Using the Results of those Calculations, We Can Define a Lognormal Distribution that is a Good Approximation to the Sugiyama-Yang Bar Graph

A Lognormal Distribution that Models Satellite Weight Growth

MCR
Critical Thinking. Solutions Delivered.

	Average Weight Growth (x)	Post-1990 Number of Units (f)	fx	fx^2
	15.5	6	93.0	1441.50
	25.5	5	127.5	3251.25
	35.5	1	35.5	1260.25
	55.5	1	55.5	3080.25
	65.5	3	196.5	12870.75
	105.5	2	211.0	22260.50
	Sums =	18	719.0	44164.5
		Approximate Mean =	39.94	μ
		Approximate Variance =	908.50	σ^2
		Approximate Standard Deviation =	30.14	σ
	Assuming a Lognormal Distribution:			
P =	3.4621458	μ =	39.94	
Q^2 =	0.4506876	σ =	30.14	
Q =	0.6713327	Mode =	20.32	





Statistics of the Lognormal Weight-Growth Distribution

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

- Using the Parameters μ and σ of the Lognormal Distribution, We Can Now Calculate All its Statistics
- If the CARD-Specified Weight for a Particular Program is 6,000 Pounds, We Can Use the Weight-Growth Information to Determine the Program's Weight Distribution
- From the Table to the Right, We See that
  - Mean Weight Growth is 39.94% of 6,000 Pounds, namely 2,396.40 Pounds
  - Standard Deviation of Weight Growth is 30.14% of 6,000 Pounds, namely 1,808.40 Pounds
- Therefore the Weight-Growth Distribution of this Program is Approximately Lognormal with μ = 2,396.40 Pounds and σ = 1,808.40 Pounds
- It Follows that the Weight itself has a Lognormal Distribution with μ = 8,396.40 Pounds and σ = 1,808.40 Pounds

Lognormal Weight-Growth Distribution (%)	
Mean	39.94
Median	31.89
Mode	20.32
Std Dev	30.14
0th Percentile	0.00
1st Percentile	6.69
5th Percentile	10.57
10th Percentile	13.49
20th Percentile	18.12
30th Percentile	22.42
40th Percentile	26.90
50th Percentile	31.89
60th Percentile	37.80
70th Percentile	45.34
80th Percentile	56.10
90th Percentile	75.38
95th Percentile	96.20
99th Percentile	152.00

Estimating Using the Weight-Growth Distribution

- Suppose $C = 47.34 + 0.67W$ is the Estimating Relationship Expressing Satellite Cost in Terms of its Weight at Authority to Proceed (ATP), where C is Denominated in $K and W in Pounds
- If W is a Random Variable Having Mean μ and Standard Deviation σ, then $0.67W$ is also a Random Variable, and it Has
  - Mean = 0.67μ
  - Standard Deviation = 0.67σ
- Furthermore, $C = 47.34 + 0.67W$ Has
  - Mean = $47.34 + 0.67\mu$
  - Standard Deviation = 0.67σ
- Finally, if the Distribution of W is Lognormal, then the Distribution of C is a "Shifted" Lognormal, Each Possible Value of which is 47.34 Larger than 0.67 Times the Corresponding Value of W

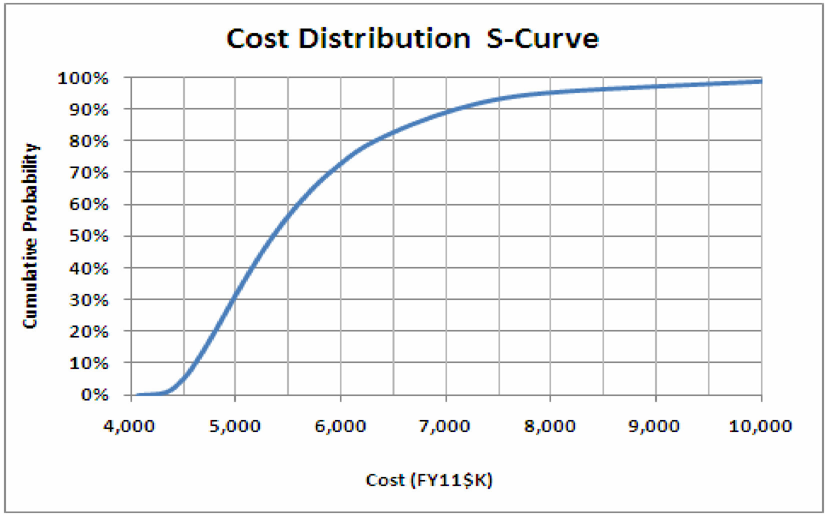
Now to Find the Cost Probability Distribution

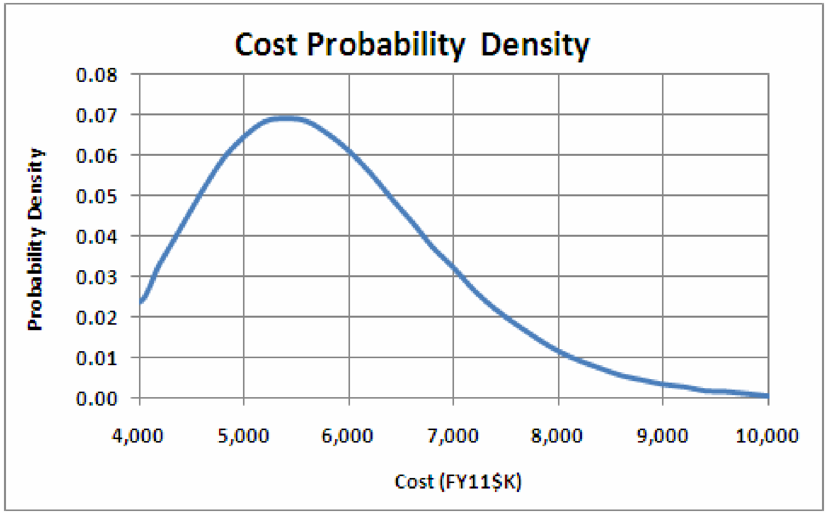
- To Obtain the Satellite Cost Distribution, We Apply the Weight-Based CER to the Entire Weight Distribution, Instead of to one Specific Value of Weight, Such as the 6,000 Pounds Listed in the CARD
- In Applying this Method, We are, from the Risk Point of View, Using *A Priori* Risk Information about the Cost Driver (Weight), Instead of *A Posteriori* Risk Information about the Cost itself
- More Precisely, We are Using Weight-Growth Information instead of Cost-Growth Information
- The Basic Idea of Doing Risk Analysis This Way is not New, but the Process that Implements it here Might be

The Lognormal Distribution of Cost, Given Weight

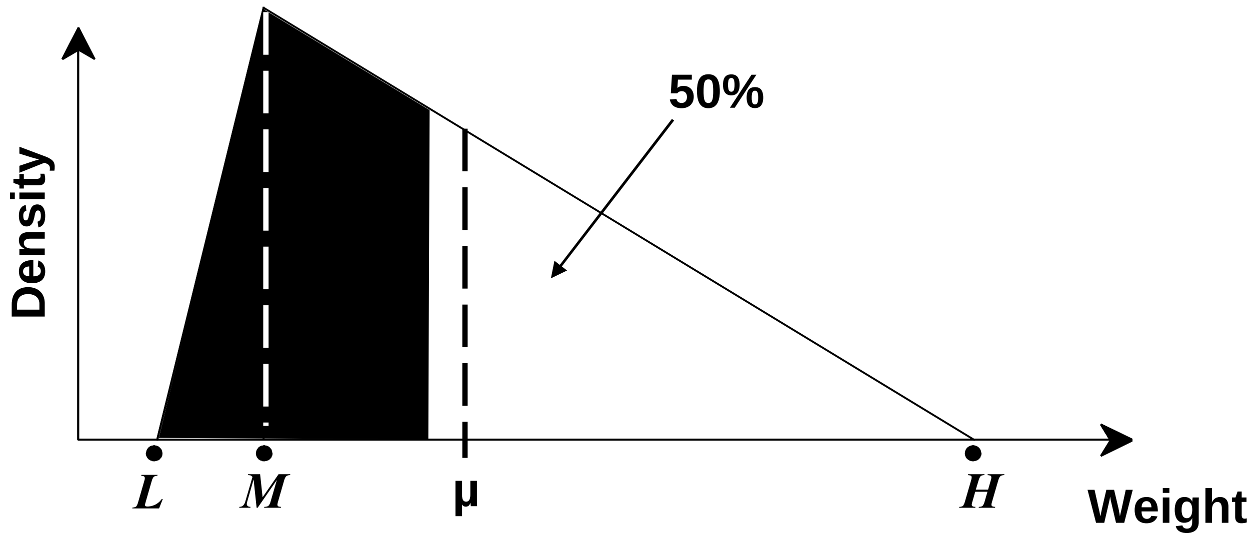
CRITICAL THINKING. SOLUTIONS DELIVERED.

Lognormal Statistics	Lognormal Weight-Growth Distribution (%)	Weight W (pounds) Lognormal Distribution	Shifted Lognormal Cost ($K) Distribution (C = 47.34 + 0.67W)
Mean	39.94	8,396.67	5,673.11
Median	31.89	7,913.12	5,349.13
Mode	20.32	7,219.02	4,884.08
Std Dev	30.14	1,808.48	1,211.68
0th Percentile	0.00	6,000.00	4,067.34
1st Percentile	6.69	6,401.31	4,336.22
5th Percentile	10.57	6,634.13	4,492.20
10th Percentile	13.49	6,809.28	4,609.56
20th Percentile	18.12	7,087.33	4,795.85
30th Percentile	22.42	7,345.40	4,968.75
40th Percentile	26.90	7,613.90	5,148.65
50th Percentile	31.89	7,913.12	5,349.13
60th Percentile	37.80	8,267.81	5,586.77
70th Percentile	45.34	8,720.41	5,890.01
80th Percentile	56.10	9,366.06	6,322.60
90th Percentile	75.38	10,522.58	7,097.47
95th Percentile	96.20	11,771.77	7,934.43
99th Percentile	152.00	15,120.17	10,177.85





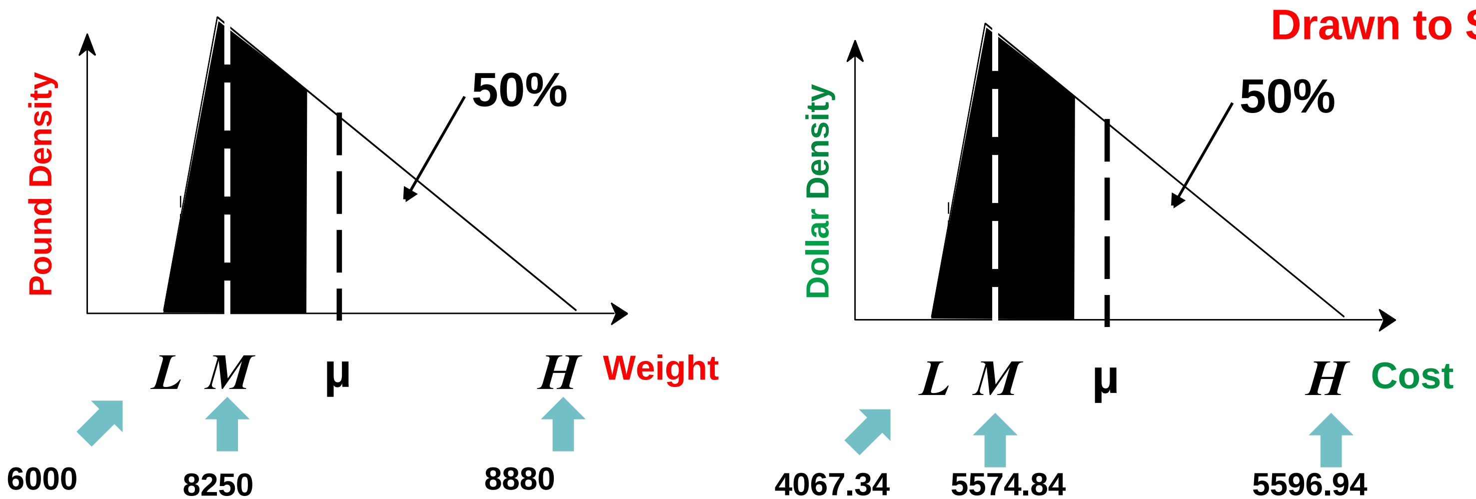
What if Weight Has a Triangular Distribution?

- The Cost Driver May, in Other Situations, Have a Triangular, Rather than a Lognormal Distribution
- In Such Situations, the Distribution is Described by a Low (optimistic) Value (L), a Most Likely Value (M), and a High (Pessimistic) Value (H)
- 

This is Another Transition from Weight to Cost

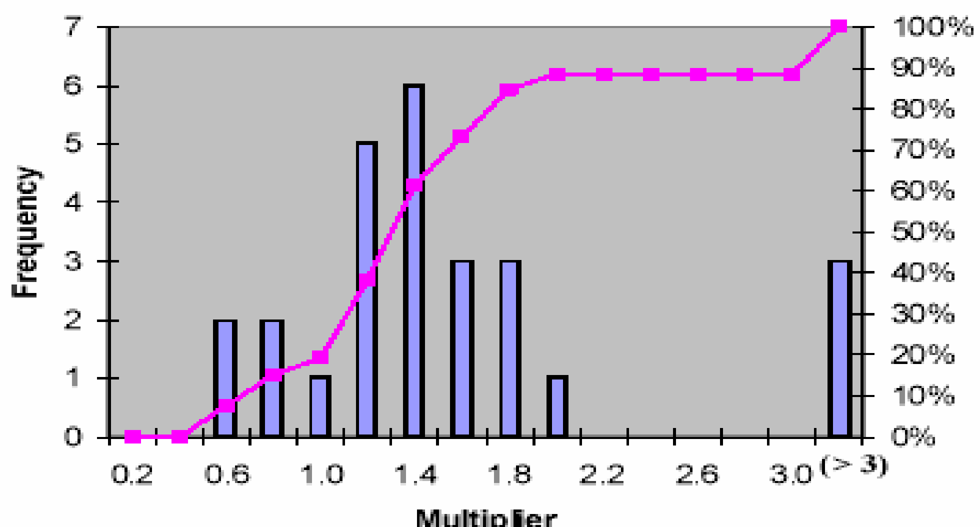
- **Suppose the Triangular Parameters are L = 6,000, M = 8,250, and H = 8,880 Pounds, Respectively**
- **Assuming the Weight-Based CER is $C = 47.34 + 0.67W$, the Cost ($K) Triangular Distribution Can be Derived from the Cost-Driver Triangular Distribution**
  - **Cost L = 47.34 + 0.67(6,000) = 4,067.34**
  - **Cost M = 47.34 + 0.67(8250) = 5,574.84**
  - **Cost H = 47.34 + 0.67(8880) = 5,596.04**

Diagrams Not Drawn to Scale



Software Lines-of-Code-Growth Historical Experience

- A 2004 Aerospace Corp. Study by J. Gayek, L. Long, K. Bell, R. Hsu, and R. Larson [Reference 5] Reported Code-Growth Data for Military Space-Related Projects, where the "Multiplier" is the Ratio of the Actual Number of Lines of Code to the Initial (e.g., CARD) Estimate
- The Following Table Summarizes the Data for 27 Projects for which Both Estimated (e.g., CARD) and Actual Number of Lines of Code were Available:



Bin intervals: $0.2(k-1) \leq m \leq 0.2k$, $k = 1, 2, \ldots, 15$, and $m > 3.0$

Approximating the Probability Distribution of "Lines of Code"

CRITICAL THINKING. SOLUTIONS DELIVERED.

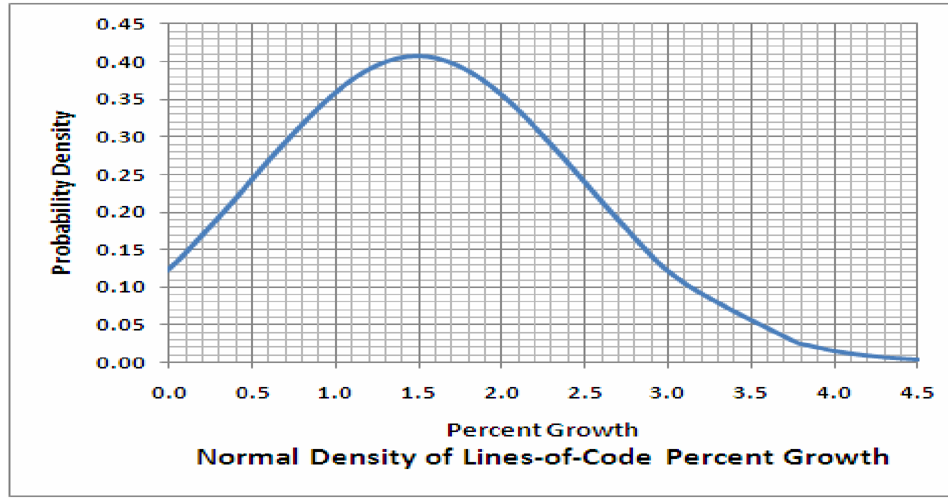
- For the 27 Projects, the Statistics were as Follows:
  - Minimum Multiplier was 0.52
  - Mean Multiplier was 1.49
  - Maximum Multiplier was 5.01
- A Glance at the Aerospace Corp. Histogram Shows that the Choice of a Probability Distribution with which to Model the Multipliers is not Obvious
  - Due to the "Spike" at the Upper End, the Triangular Distribution Does not Seem Appropriate
  - But the Spike Represents not Just One Number but Rather All Numbers between 3 and 5, so it is Reasonable to Assume that it Should be Spread Out over a Longer Interval
  - Otherwise, the Distribution Appears Fairly Symmetric about its Mode
  - Therefore, the Normal Could Very Well Serve as an Appropriate Model
- Let's Go with That and See what Happens

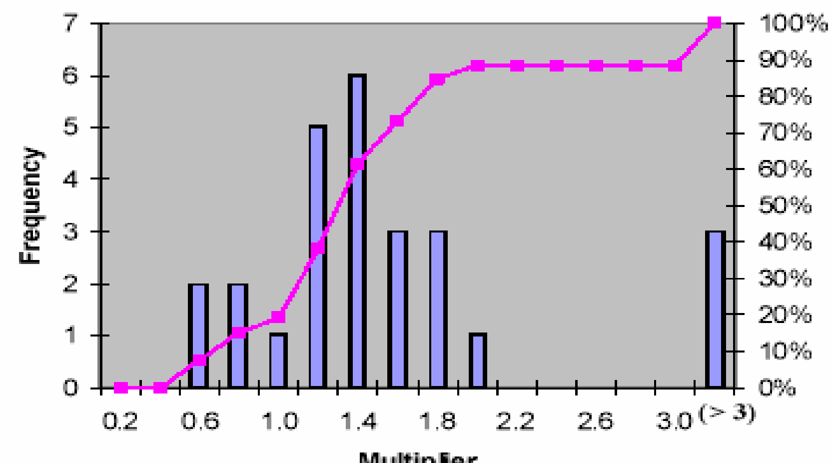
There are Additional Studies of Lines-of-Code Growth

- For our Example, We Will Use the Data from the 2003 Aerospace Corp. Study and Model them with a Lognormal Distribution
- However, When Applying this Process to a "Real" Estimate, Other Sources Focusing on Different Aspects of Code Growth Should be Consulted for Data (and CERs) Possibly More Relevant to your Specific Context
- Four such Additional Sources are the Following:
  - B. Holchin, "Code Growth Methodology," 2006 [Reference 6]
  - R. Jensen, "Estimating Software Growth," USAF Software Technology Support Center, Hill AFB UT, Space Systems Cost Analysis Group, 2008 [Reference 7]
  - R. Jones and P. Hardin, "Software Code Growth: A New Approach Based on Historical Analysis of Actuals," Navy/USMC Cost Analysis Symposium, 2007 [Reference 8]
  - M. Ross, "Software Size Uncertainty: The Effects of Growth and Estimation Variability," r2 Estimating, 2007 [Reference 10]

A Normal Distribution that Models Lines-of-Code Growth

	Representative Lines-of-Code Growth Rate	No. of Units (f)	fx	fx^2
	0.6	2	1.2	0.72
	0.8	3	2.4	1.92
	1.0	1	1.0	1.00
	1.2	5	6.0	7.20
	1.4	6	8.4	11.76
	1.6	3	4.8	7.68
	1.8	3	5.4	9.72
	2.0	1	2.0	4.00
	3.0	1	3.0	9.00
	4.0	1	4.0	16.00
	5.0	1	5.0	25.00
	Sums =	27	43.2	94
		Mean =	1.49	P
		Approximate Variance =	0.96	Q^2
Approximate Standard Deviation =			0.98	Q





Bin intervals: $0.2(k\text{-}1) \le m \le 0.2k$, $k = 1, 2, \ldots, 15$, and $m > 3.0$

Lines-of-Code Growth Normal Distribution Statistics

MCR CRITICAL THINKING. SOLUTIONS DELIVERED.

- We Use the P and Q Parameters of the Normal Distribution to Calculate All its Statistics
- How Can We Use This Information to Determine a Program's Lines-of-Code Probability Distribution?
  - Suppose the CARD-Specified New Lines of Code for Military Ground Support Software on a Particular Program is 250,000
  - From the Normal Statistics Table,
    - The Mean of the SLOC Distribution Should be 149% of 250,000 Lines, namely 372,500 Lines
    - Standard Deviation of the SLOC Distribution Should be Growth Should be 98% of 250,000 Lines, Namely 245,000 Lines
- It Follows that the Probability Distribution of SLOC for this Program is Approximately Normal with Mean = 372,500 Lines and Standard Deviation 245,000 Lines

Lines-of-Code Growth Normal Distribution (SLOC Multiplier)	
Mean	1.49
Median	1.49
Mode	1.49
Std Dev	0.98
1st Percentile	-0.79
5th Percentile	-0.12
10th Percentile	0.24
20th Percentile	0.67
30th Percentile	0.98
40th Percentile	1.24
50th Percentile	1.49
60th Percentile	1.74
70th Percentile	2.00
80th Percentile	2.31
90th Percentile	2.74
95th Percentile	3.10
99th Percentile	3.77

Estimating Using the Code-Growth Distribution

- **The Aerospace Corp. Report Offers a CER for Estimating Military Ground Support Software – It is Expressed in "Developer-Months" (DM) Rather than Dollars, so Inflation Issues are Obviated**
- **The CER is**

$$DM = 23.48 + 0.0003051 \times SLOC^{1.256}$$

- **Because this CER is not Linear, the Probability Distribution of DM is not Normal**
  - Although the Algebraic Expression for the Probability Distribution Can be Found Using Calculus, a Simpler Approach is to Directly Calculate its Percentiles
  - For Example, because 2.31 is the 80^{th} Percentile of the Multiplier Distribution, $2.31 \times 250{,}000 =$ 575,500 Lines is the 80^{th} Percentile of the SLOC Distribution
  - Therefore, the 80^{th} Percentile of the DM Distribution is $23.48 + 0.0003051 \times (575{,}500)^{1.256} =$ 5,260.2 Developer-Months
  - This is 438.3 Developer-Years, e.g., 109.6 FTE over Four Years

The Normal Distribution of DM, Given SLOC

- **Summary: To Obtain the DM Distribution, We Applied a SLOC-Based CER to the Entire SLOC Distribution, Instead of to a Specific Possible Value of SLOC, like 250,000**
- **Again, from the Risk Point of View, We Used Risk Information about the DM Driver (SLOC), Instead of about the DM itself**

Lines-of-Code Growth		
Normal Distribution		DM Distribution
(SLOC Multiplier)		$DM = 23.48+0.0003051 \times SLOC^{1.256}$
Median	1.49	3055.81
10th Percentile	0.24	323.71
20th Percentile	0.67	1127.85
30th Percentile	0.98	1808.21
40th Percentile	1.24	2436.42
50th Percentile	1.49	3055.81
60th Percentile	1.74	3702.26
70th Percentile	2.00	4420.53
80th Percentile	2.31	5292.49
90th Percentile	2.74	6551.70
95th Percentile	3.10	7630.87
99th Percentile	3.77	9740.11
NOTE: SLOC = Multiplier $\times$ 250,000		

CRITICAL THINKING. SOLUTIONS DELIVERED.

What's Really Going On Here?

- **There are Two Basic Ways of Using Risk and Uncertainty Information to Establish Probability Distributions of Cost**
  - List All the Risk Issues Facing the Program (some may be correlated) and Attempt to Assess their Combined Impact on Program Cost
  - Determine the Major Cost Driver (or Drivers) of Program Cost and Attempt to Assess the Impact of Risks on Them (the drivers)
- **We're Discussing Another Approach: If We Have Faith that the Estimating Relationship is Valid, the Impact of Cost-Driver Risk on Cost Can be Algebraically Calculated**

Contents

- Current Status of Estimating Relationships
- Why Cost and Schedule Must be Modeled as Probability Distributions
- Why Cost and Schedule Drivers Must be Modeled as Probability Distributions
- What's Wrong with the Way Cost-on-Cost Estimating Relationships are Modeled and How to Fix It
- Applications
- Summary

What are Cost-on-Cost CERs?

- Cost-on-Cost (CoC) CERs Estimate BTL or "Level-of-Effort" (LOE) Costs
- BTL or LOE Costs are Costs of WBS Elements that do not Represent Actual Products but rather Management and Engineering Activities that Support Development and Production of those Products
- Among BTL or LOE Costs are Included those of the Following Activities, for Example . . .
  - System Engineering (SE)
  - Program Management (PM)
  - Component and System Integration and Assembly
  - Component and System Testing
  - Documentation
  - Production Engineering
  - Operator Training
  - Site Acceptance and Testing
  - Installation and Checkout

The Simplistic Approach to Cost-on-Cost CERs

- **CoC CERs are Usually Expressed as Factor CERS, i.e., Corresponding to Each Program's BTL Cost, so there is a Factor that we Multiply Times the Cost of the "Prime Mission Product" (PMP), namely the Total Cost of the Program's Hardware and Software**
- **The Factor Has Traditionally Been Expressed as a Decimal or Percentage Greater than Zero**
  - **For Example, SE Cost Might be Estimated as 60% of PMP Cost, i.e. SE Cost = 0.60 × PMP Cost.**
  - **PM Cost Might be Estimated as 114% of PMP Cost, i.e., PM Cost = 1.14 × PMP Cost.**
- **A Factor is Selected Based on Several Considerations**
  - **The Contractor's *Modus Operandi* and Accounting System**
  - **The Kind of Work Planned, e.g., Hardware-Intensive vs. Software-Intensive**
  - **The Planned Schedule**

The Problem with The Simplistic Approach

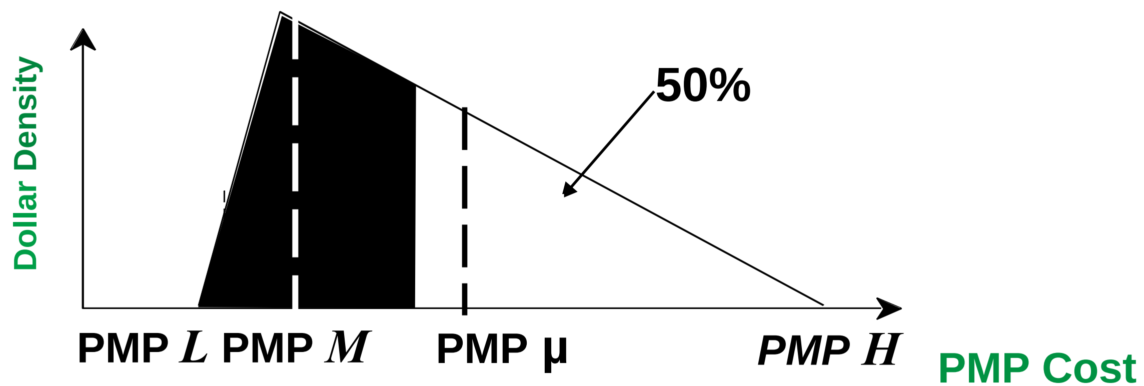
- **CoC CERs are Actually Misnamed – They Really Should be Called "Cost-Estimate-on-Cost-Estimate" CERs Because We Use PMP Cost Estimates (<u>not Actuals</u>) to Estimate BLT and LOE Costs**
- **However, Values of Most (if not all) Drivers of PMP Costs, and therefore PMP Costs Themselves, are not Precisely Known at the Estimating Stage, but are Subject to Change due to Various Uncertainties, e.g., Risks, Schedule Slippage, Requirements Changes**
- **It Follows that PMP Cost is a Probability Distribution, not a Number, so the Simplistic Approach Lacks not Only Logic, but also Credibility**
- **Fortunately, This Problem Can be Circumvented by Applying the Idea Behind Covert's EIV Estimating Process**

One Way to Apply the Errors-in-Variables Concept

- **EIV CERs Recognize the Fact that Cost-Driver Values are not Hard Numbers at the Estimating Stage**
- **Therefore Drivers of EIV CERs Must be Used in the Form of Probability Distributions**
  - **A CER for SE Cost Must Use the PMP Cost Probability Distribution, not Just the Point Estimate, as a Driver**
  - **This is True Whether or not the CER is a Factor CER or One of Some Other Algebraic Form**
- **It Follows that SE Cost and Other BTL Costs are also Probability Distributions, Ones Algebraically Derived from the PMP Probability Distribution**
- **Finally, All of this Means that Total Program Cost is a Statistical Compilation of Several Probability Distributions: PMP, SE, Other BTL Cost, Maybe Some Additional Cost Probability Distributions**

It Gets Worse (Actually Just Slightly More Complicated)

- **Suppose SE Cost is Estimated Using a Factor CER with PMP Cost as its Driver, so that SE Cost = Factor × PMP Cost**
- **We Already Know that PMP Cost is a Probability Distribution such as**



- **The Factor, as Noted Earlier, is Typically Expressed as a Percentage - However, the Factor Itself is also a Probability Distribution, because we Do not Know this Percentage Precisely**

How Do We Find the Factor's Probability Distribution?

- We Might Have Historical Information on BTL Costs of Programs of the Kind Being Estimated
- If We Lack Information on Programs of that Specific Type, We Can Apply the Experience-Based Summary Information Provided by S.M. Allard in his Report "Cost Factors for Estimating and Analysis" [Reference 1]
- If We Lack Information that Would Point Us to a Particular Probability Distribution, the Triangular is Easy to Use and Quite Adequate
- To Apply a Triangular Distribution, We Need Only…
  - A Low (optimistic, i.e., low-cost) Percentage L
  - A Most Likely Percentage M
  - A High (pessimistic, i.e., high-cost) Percentage H

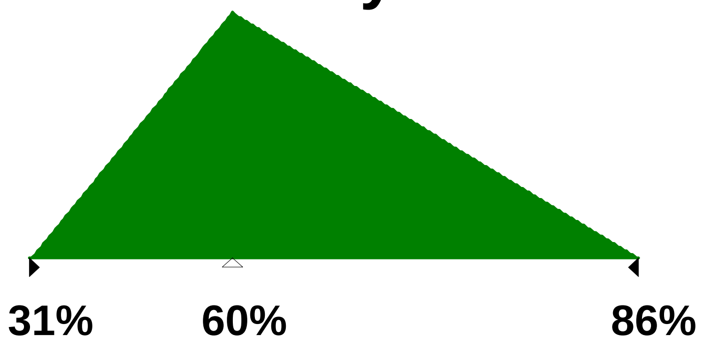
Allard's Table of Factors for Development Phase

FAA ASD-410 Pocket Estimating Guide		*** Coordinate with ASD-410 before use ***				12 Dec 01 Ver 0.33
WBS	***DEVELOPMENT***	**Low Tendency**	**Low**	**ML**	**High**	**High Tendency**
3.2	System Engineering	Hardware Intensive	31%	**60%**	86%	Software Intensive
3.3.1.2	*Hardware less NRE, AUC	All COTS	100%	**150%**	200%	New Development
3.3.3	HW/SW Integ., Ass'y, Test & Chkout	Hardware Intensive	10%	**16%**	24%	Software Intensive
3.4.1	Facility Planning & Design	Software Intensive	2%	**24%**	47%	Hardware Intensive
3.5.1	System Dvlpmt. Test & Eval.	Minor Modification	5%	**15%**	27%	New Capability
3.6	Documentation	Minor Modification	1%	**21%**	27%	New Capability
3.7.3	Support & Hdlg Equip. Acq. (CSE)	Minor Modification	2%	**8%**	11%	New Capability
3.7.4	Support Fac. Const. / Conv. / Exp.	Software Intensive	10%	**14%**	20%	Hardware Intensive
3.7.5	Support Equip. Acq. / Mod. (PSE)	Minor Modification	1%	**10%**	34%	New Capability
3.7.7	Initial Spares & Repair Parts Acq.	Software Intensive	1%	**19%**	39%	Hardware Intensive
3.7.8	Initial Training	Minor Modification	1%	**10%**	17%	New Capability
Factors applied to sum of (WBS 3.3.1 Hdw + WBS 3.3.2 SW) with exception of WBS 3.3.1.2						
*Factor applied to WBS 3.3.5, Production, Average Unit Cost (AUC)						

- **Focusing on the SE Factor, Notice that the Kind of Program (hardware-intensive vs. software-intensive) Determines the Relative Likelihood of the "Low" and "High" Values**

Applying Information from Allard's Table of Factors

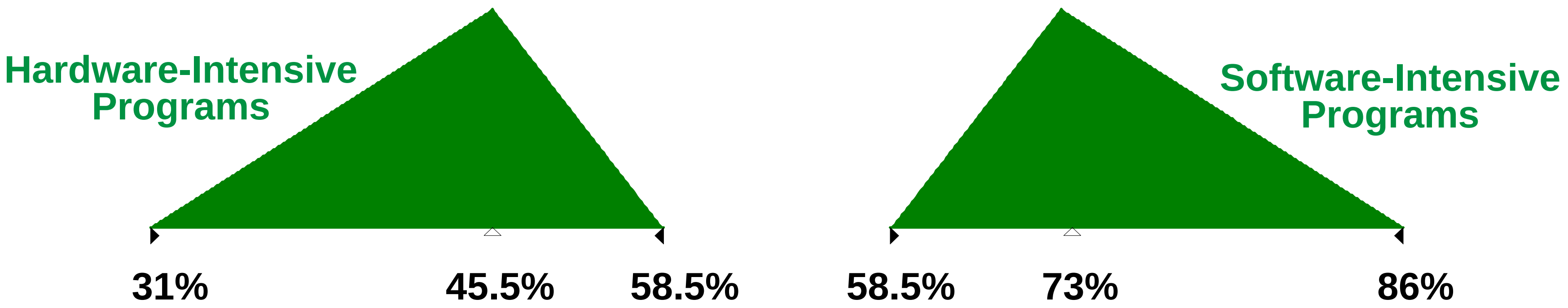
- **If We were to Use the SE Factors Directly from the Table, We would Assume the Triangular Distribution Below for the SE Probability Distribution**



- **However, This is not Proper Use of the Data in the Table – We Have Failed to Take Account of the Differing Low and High "Tendencies"**
  - **SE Factor for Hardware-Intensive Programs is Closer to 31%**
  - **SE Factor for Software-Intensive Programs is Closer to 86%**

How Can We Take the Kind of Program into Account?

- If We Have no More Detailed Knowledge of SE Factors than that already Included in the Table, We Have to Make Inferences (i.e., assumptions)
- One Such Assumption Could be the Following:
  - For Hardware-Intensive Programs, Assume L = 31%, M = (31%+60%)/2 = 45.5%, and H = (31%+86%)/2 = 58.5%



  - For Software-Intensive Programs, Assume H = 86%, M = (60%+86%)/2 = 73%, and L = (31%+86%)/2 = 58.5%
- We Will Make that Assumption Throughout this Presentation

Contents

- Current Status of Estimating Relationships
- Why Cost and Schedule Must be Modeled as Probability Distributions
- Why Cost and Schedule Drivers Must be Modeled as Probability Distributions
- What's Wrong with the Way Cost-on-Cost Estimating Relationships are Modeled and How to Fix It
- Applications
- Summary

Schedule as a Cost Driver

- Schedule = Project Duration in Number of Months
- Run-Out Rate = Project Expenditure in Dollars per Month
- Cost = Schedule × Run-Out Rate
- Suppose Initial Estimates are as Follows:
  - Schedule = 48 Months
  - Run-Out Rate = $80,000 per Month
- Naïve Estimate of Project Cost

$$C = (\text{Number of Months}) \times (\text{Dollars per Month})$$
$$= 48 \times \$80{,}000 = \$3{,}840{,}000$$

- We Call this Estimate "Naïve," because it Doesn't Take into Consideration Schedule and Run-Out-Rate Uncertainties that May Lead to Increased Costs

We Are Not Yet Done with This Estimate

- This Naïve Estimate Fails to Account for Schedule and Run-Out-Rate Uncertainties that May Lead to Increased, or at Least Different, Costs
- Numerical Values of the Two Cost Drivers are in Fact Being Estimated at the Same Time as the Costs are Estimated
- Critical Cost-Driver Uncertainties are the Following:
  - Schedule Slippage Beyond the Originally Planned 48 Months
  - Uncertainty in the Labor Rates, Staffing Mix, and Technical Difficulty that May Increase, or at Least Change, the Run-Out Rate from that Originally Estimated
- We Can Account for the Impact of these Risk Issues by Considering Each Cost Driver to be a Random Variable with a Probability Distribution, such as …
  - Lognormal Schedule with μ = 48 Months, σ = 6 Months
  - Triangular Run-Out Rate with L = \$72,000, M = \$80,000, and H = \$100,000

Simulating Schedule, Run-Out Rate, and Cost Distributions

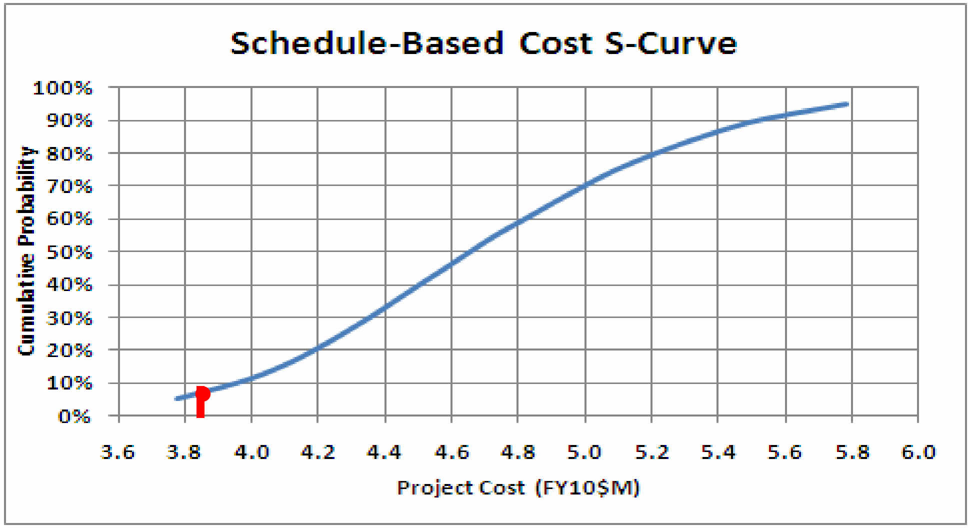
Trial Number	Schedule S (Months)	Run-Out Rate R (Dollars/Month)	Project Cost S×R (Dollars)
1	52.42	86,284.45	4,522,774.26
2	60.24	80,429.07	4,844,736.58
3	55.85	75,565.29	4,220,508.05
4	55.21	96,082.80	5,304,712.30
5	62.70	81,219.77	5,092,588.42
6	58.29	85,214.80	4,966,970.45
7	47.92	94,030.24	4,506,093.04
8	50.53	77,725.59	3,927,185.85
•	•	•	•
•	•	•	•
•	•	•	•
9994	51.56	78,285.95	4,036,254.98
9995	49.73	82,929.30	4,124,334.10
9996	74.14	82,766.14	6,136,035.65
9997	53.81	79,144.13	4,258,373.74
9998	56.22	87,691.40	4,930,217.53
9999	61.84	98,028.08	6,062,413.45
10000	48.05	76,641.05	3,682,922.98

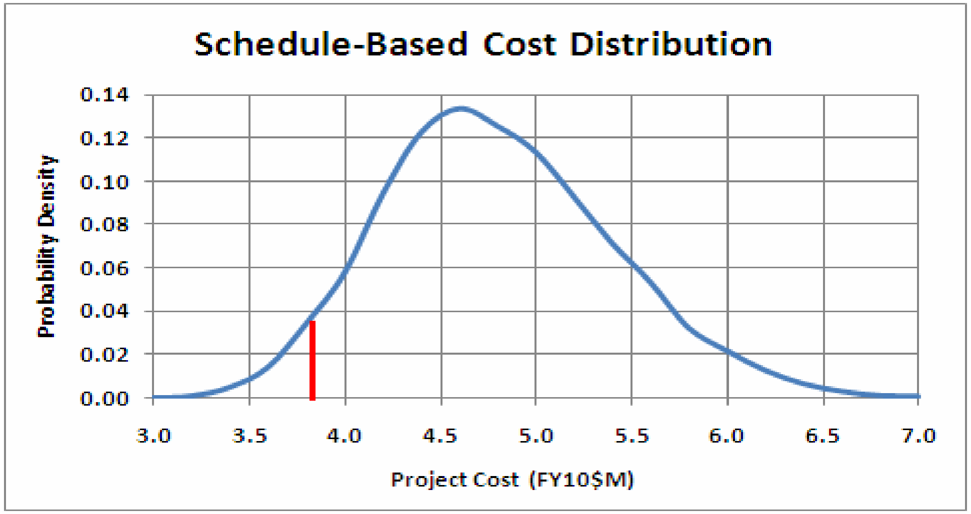
Note: Project Cost = Schedule × Run-Out Rate

Statistics of the Schedule-Based Project Cost Estimate

Mean =	699.20
Std Dev =	153.59

SE Cost Percentiles		
	5th	3,776,496.29
(Naïve Estimate)	6.66th	3,840,000.00
	10th	3,960,647.96
	15th	4,090,386.92
	20th	4,188,628.15
	25th	4,274,450.09
	30th	4,355,567.62
	35th	4,430,185.35
	40th	4,504,841.64
	45th	4,580,936.71
	50th	4,655,239.39
	55th	4,731,781.40
	60th	4,819,804.64
	65th	4,906,121.82
	70th	4,995,658.22
	75th	5,092,588.42
	80th	5,212,382.53
	85th	5,348,809.76
	90th	5,514,877.42
	95th	5,781,145.29





A Software Costing Problem

- Cost = Hours×(Dollars per Hour)
- The Following Numbers Characterize the Drivers of Software Cost (*C*)
  - Number of Source Lines of Code (*SLOC*)
  - Number of Developer Hours (*HRS*)
  - Developer Labor Rates (*$/HR*)
  - Developer Productivity (*SLOC/Month*)
- Based on those Quantities, a CER Proposed for the Cost of a Software Project is

$$C = [(SLOC \times (\$/HR) \times (HRS/Month)] \div (SLOC/Month)$$

- ("Units Check") A Little Arithmetic Shows that

  C = [(SLOC×$/HR×(HRS/Month)]×(Months/SLOC),

  where the SLOC and "Month" Terms Cancel Out, Leaving C = ($/HR)×HRS, namely, "Dollars"

A Naïve Estimate of Software Cost

- The CARD or Other Project Description Specifies that SLOC = 200,000
- Economic and Historical Experience Shows that
  - (Mean) $/HR = 150
  - (Mean) SLOC/Month = 139
- We Assume as a Fact that HRS/Month = 168
- Then the CER Discussed on the Previous Chart Estimates Software Project Cost to be

$$C = [(SLOC \times \$/HR \times 168] \div (SLOC/Month)$$

- For this Project, it then Follows that

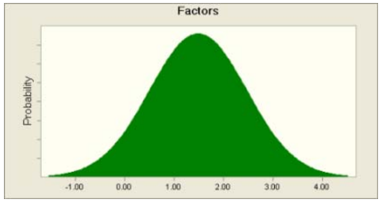
$$C = 200{,}000 \times 150 \times 168 \div 139 = \$36{,}258{,}992$$

We Are Not Yet Done with This Estimate

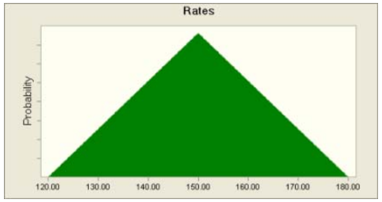
- As it Stands, We Call this Estimate "Naïve," because it Fails to Account for Cost-Driver Uncertainties that May Lead to Increased Costs
- Numerical Values of the Cost Drivers are Estimated at the Same Time as the Costs are Estimated
- Major Cost-Driver Uncertainties are the Following:
  - SLOC Growth Beyond the CARD-Specified 200,000 Lines
  - Uncertainty in the Mean Labor Rate due to Possible Staff Realignments as the Project Proceeds to Completion
  - Uncertainty in Developer Productivity due to Probable Mis(under)estimation of Project Difficulty
- We Can Account for the Impact of these Risk Issues by Considering Each Cost Driver to be a Random Variable with a Probability Distribution
- The Estimate is then Expressed, not as a Definite Number of Dollars, but as a Probability Distribution of Project Cost that Includes the Impact of Risk

Cost-Driver Probability Distributions

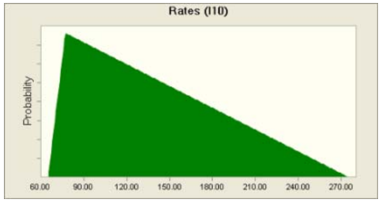
- Based on Historical SLOC Growth Experience (which appears to be a multiple with mean 1.49 and standard deviation 0.98), We Model SLOC as a Normal Distribution with Mean 298,000 and Standard Deviation 196,000



- Economic Studies (of the kind conducted by the Bureau of Labor Statistics) Indicate that \$/HR has a Triangular Distribution with L = 120, M = 150, and H = 180



- From Aerospace Corp. [Reference 5] and D. Reifer [Reference 9] Studies, We Infer that Developer Productivity in SLOC/Month is Reasonably Modeled as Triangular with L = 65, M = 77, and H = 275



Monte Carlo Simulation of the Software Cost Distribution

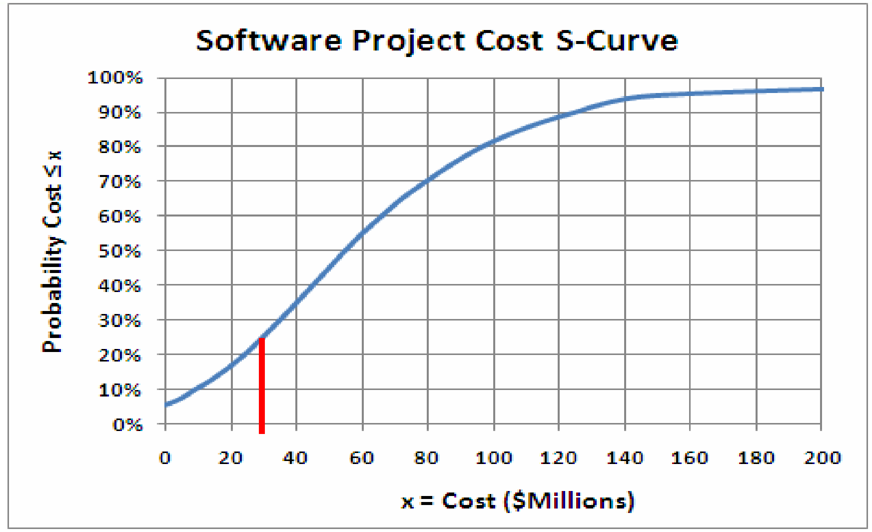
Trial values	Risk Multiple	Code Growth	Labor Rates	Productivity	Spec SLOC	HRS/Month	Project Cost ($)
1	4.353113971	3.343265262	145.620217	111.838793	200,000	168	146,264,629.42
2	2.132871738	2.167614083	162.759199	165.410384	200,000	168	71,664,490.38
3	1.858745357	1.80101806	158.993596	154.055711	200,000	168	62,453,843.99
4	1.494854919	1.186994153	144.813019	114.989224	200,000	168	50,227,125.27
5	1.709847333	2.102261555	136.451013	167.76686	200,000	168	57,450,870.40
6	2.247484217	2.036110547	137.626725	124.683068	200,000	168	75,515,469.69
7	-0.557033384	-0.67834929	147.375707	179.472558	200,000	168	-18,716,321.71
8	0.17085787	0.24026301	163.350774	229.706415	200,000	168	5,740,824.43
•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•
9994	0.998464749	1.260687257	145.276381	183.429693	200,000	168	33,548,415.57
9995	0.389029151	0.589649643	146.01486	221.314033	200,000	168	13,071,379.48
9996	1.58746048	2.313262625	155.279758	226.275151	200,000	168	53,338,672.12
9997	0.424355663	0.503648065	138.930014	164.889594	200,000	168	14,258,350.27
9998	0.667457352	1.068242153	128.920709	206.333086	200,000	168	22,426,567.04
9999	-0.061275713	-0.042208624	158.857655	109.426112	200,000	168	-2,058,863.96
10000	2.343028882	1.194585188	169.010381	86.1693592	200,000	168	78,725,770.42

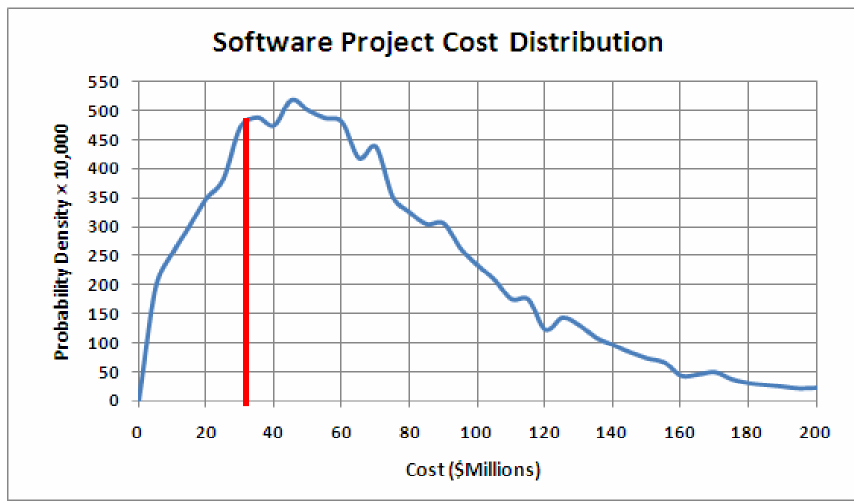
Note: Risk Multiple = Code Growth Multiple × Labor Rate ÷ Productivity

The Software Cost Estimate and its Statistics

Risk Multiple Statistics	
Sample Runs	10,000
Mean	1.81
Median	1.61
Standard Deviation	1.41
Variance	1.99
Skewness	0.7690
Kurtosis	4.30
Coeff. of Variability	0.7791
Minimum	-4.43
Maximum	9.21
Range Width	13.64
Mean Std. Error	0.01

	Percentiles	Project Cost
	10%	8,406,901.80
	20%	23,459,989.97
	30%	34,256,754.02
(Naïve Estimate)	31.94%	36,258,992.00
	40%	44,242,696.21
	50%	54,088,494.94
	60%	65,212,844.02
	70%	78,655,131.19
	80%	95,630,744.90
	90%	123,846,259.38
	95%	148,506,391.72



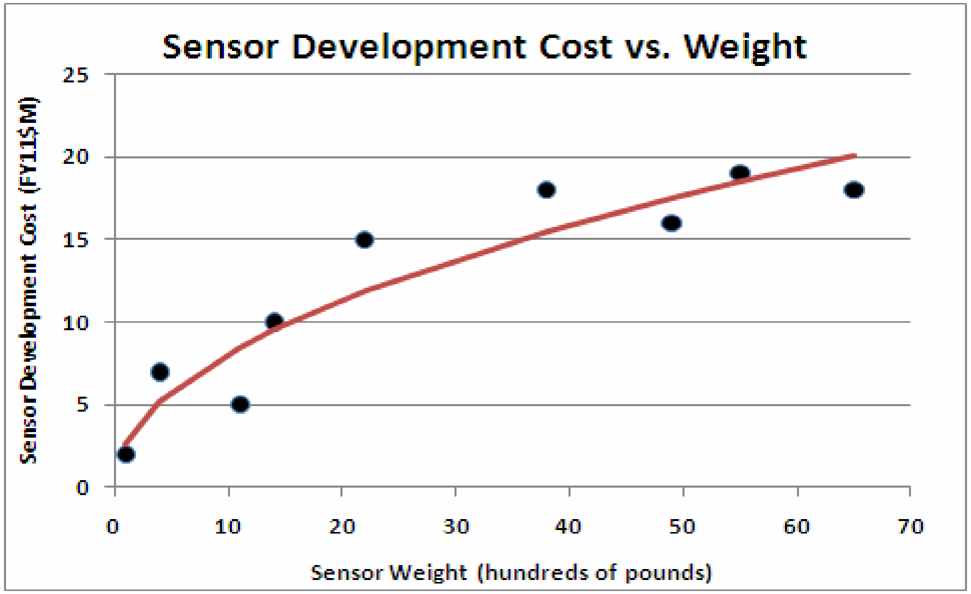


A Hardware Costing Problem

- **Historical Data on Sensor Development Cost vs. Sensor Weight Appears in the Table to the Right**
- **Applying the Minimum-Percentage-Error, Zero-Percentage-Bias (MPE/ZPB or "ZMPE") CER-Derivation Technique Yields the CER**

$$C = 2.65W^{0.48}$$

Weight (hundreds of pounds)	Development Cost (FY11$M)
55	19
38	18
14	10
4	7
49	16
65	18
22	15
1	2
11	5



A Naïve Estimate of Sensor Development Cost

- The CARD or Project Description Specifies that Sensor Weight Will be 5,200 Pounds
- That Means that the Cost Driver Value to be Input into the CER on the Previous Chart (weight expressed in hundreds of pounds) is 52
- Applying the CER, We then Estimate the Sensor's Development Cost to be

$$C = 2.65 \times (52)^{0.48} = 2.65 \times 6.66 = 17.65$$

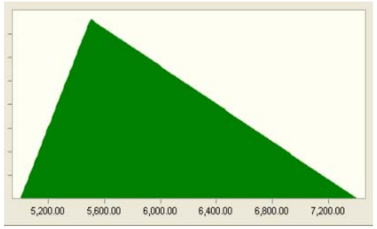
  Millions of FY11 Dollars
- This CER has the Following Quality Metrics
  - Percentage Standard Error = 25.82%
  - Percentage Bias = 0.00%
  - R^2 (between estimates and actuals) = 88.04%

We Are Not Yet Done with This Estimate

- As it Stands, this Estimate is "Naïve," because it Fails to Account for Cost-Driver Uncertainties that May Lead to Increased Costs
- We're Really Estimating the Weight at the Same Time as We're Estimating the Cost
- Major Cost-Driver Uncertainties are the Following:
  - Sensor Weight Growth Beyond the CARD-Specified 5,200 Pounds
  - Statistical Uncertainty in the CER as Measured by the Percentage Standard Error
- We Can Account for the Impact of these Risk Issues by Considering
  - Sensor Weight as a Random Variable with a Probability Distribution
  - The CER Standard Error, which Introduces Additional Uncertainty
- The Estimate is then Expressed, not as a Definite Number of Dollars, but as a Probability Distribution of Cost that Includes the Impact of these Uncertainties

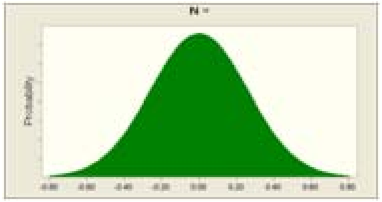
Cost-Driver Probability Distributions

- We Understand that Actual Sensor Weight is not 100% Certain to be the CARD-Specified Value of 5,200 Pounds
- Suppose Engineering Assessment of the Particular Task, Combined with Historical Experience in Sensor Development Leads Us to Believe the Weight is Better Expressed as a Triangular Probability Distribution with L = 5,000 Pounds, M = 5,500 Pounds, and H = 7,400



- Without Taking Account of this Uncertainty, We (and others) Will be Encouraged to Place More Confidence in the Estimate than is Warranted

What About the Statistical Error of the CER?



- The CER has Standard Error of 25.82%, a "One-Sigma" Value
- We Can Put this in Context by Modeling the CER Error as a Normally Distributed Random Variable with Mean 0% (because the bias is 0.00%) and Standard Deviation 25.82%
- Try this …
  - Draw a Random Number (call it *w*) from the Weight Triangular Distribution and Run it Through the CER to Obtain a Possible Value of the CER-Based Estimate (call it *c*)
  - Next Draw a Random Number (call it ε) from the CER-Error Normal Distribution Having Mean 0 and Standard Deviation 0.2582
  - Then Multiply c by 1+ε to Obtain one Possible Value of the Cost
  - The Set of All c(1+ε) Values Constitute the Cost Probability Distribution

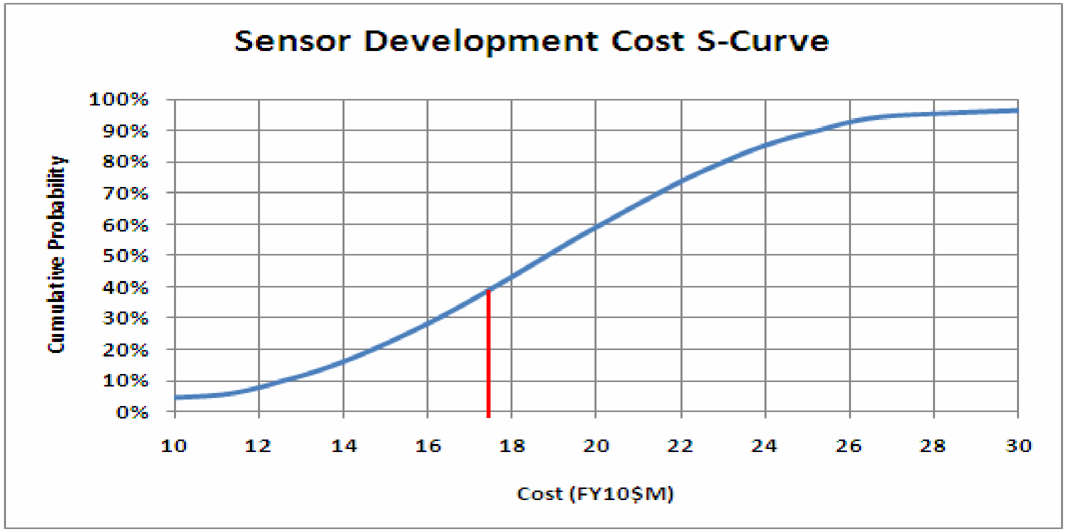
Monte Carlo Simulation of the Sensor Cost Distribution

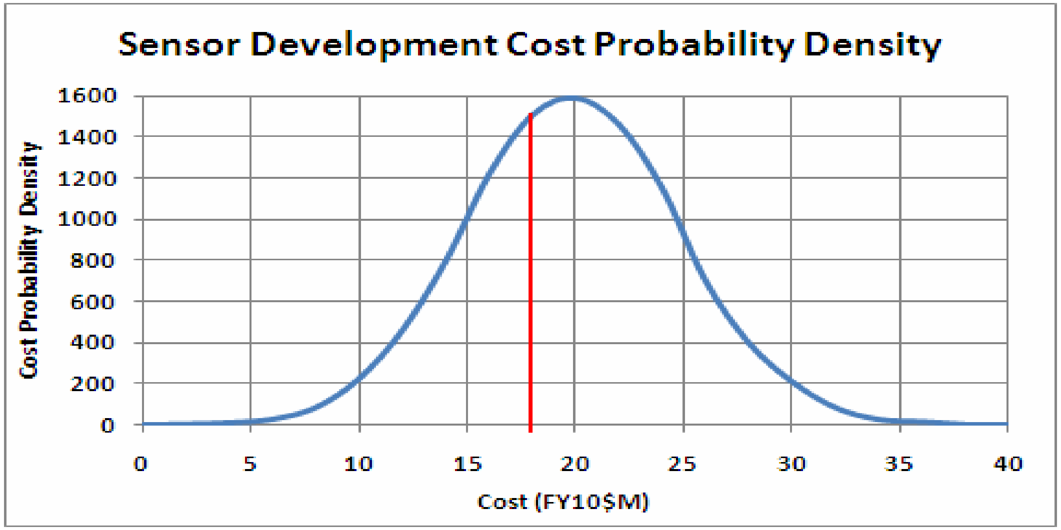
Random Number Trial	Random Percentage Error (ε)	Random Sensor Weight (*w*)	CER-Based Estimate $2.65w^{0.48}$	Cost Distribution Value w(1+ε)
1	-0.09	7,034.44	20.41	18.58
2	0.18	5,690.37	18.44	21.77
3	0.14	6,527.86	19.69	22.52
4	0.18	6,092.79	19.05	22.47
5	-0.08	6,132.37	19.11	17.60
6	-0.08	5,666.06	18.40	17.00
7	-0.27	6,300.71	19.36	14.16
8	0.15	6,268.28	19.31	22.15
•	•	•	•	•
•	•	•	•	•
•	•	•	•	•
9994	-0.22	6,138.25	19.12	14.93
9995	0.21	5,219.71	17.69	21.34
9996	-0.47	5,820.88	18.64	9.81
9997	0.06	7,104.24	20.51	21.76
9998	-0.34	5,090.31	17.48	11.50
9999	-0.24	5,666.71	18.40	14.05
10000	0.11	6,956.19	20.30	22.52

The Sensor Cost Estimate and its Statistics

Mean =	18.85
Std Dev =	4.96

Sensor Cost Percentiles	
5th	10.72
10th	12.56
15th	13.76
20th	14.68
25th	15.48
30th	16.23
35th	16.91
40th	17.56
(Naïve Estimate) 40.65th	17.65
45th	18.19
50th	19.43
60th	20.09
65th	20.75
70th	21.44
75th	22.16
80th	22.99
85th	23.90
90th	25.19
95th	27.07



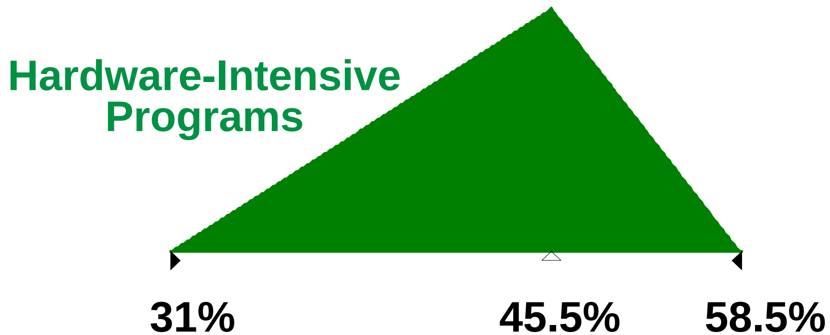


System Engineering Costs

- As Discussed Earlier in this Presentation, SE Costs are Typically Estimated as a Percentage (might be greater than 100%) of PMP Costs, i.e. by Using a Factor CER of the Form

$$SE\ Cost = a \times PMP\ Cost$$

- For "Hardware-Intensive" Projects, We Can Apply the Conclusions of the Allard Study to Model the Percentage Factor Using a Triangular Distribution with a "Most Likely" Value of 45.5%



A Naïve Estimate of System Engineering Cost

- Suppose, to Estimate the PMP Cost of a Cooling System (Hardware-Intensive) for a Satellite, We Use the CER

$$PMP\ Cost = 10.61 + 27.34W^{0.86}$$

where W is the System Weight, say 78 Pounds, as Listed in the CARD

- A Naïve Estimate of the Cooling System PMP Cost is then $10.61+27.34\times78^{0.86}$ = \$1,169.39K
- Noting that the Most Likely SE Factor for a Hardware-Intensive Project is 45.5%, SE Cost Would be Estimated as 45.5% of 1,169.39, namely \$532.07K

We Are Not Yet Done with This Estimate

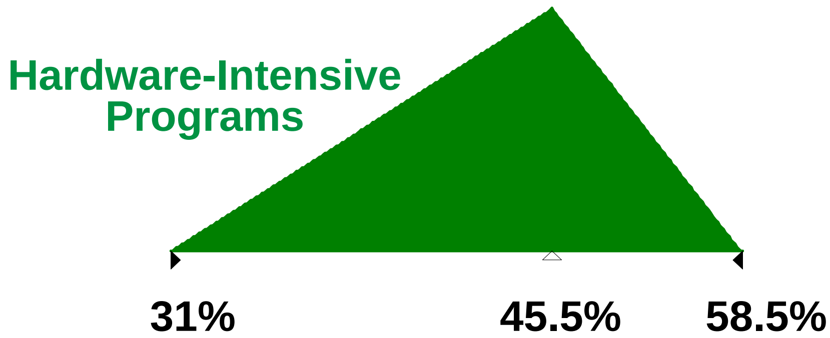
- This Estimate is “Naïve,” because it Fails to Account for the Facts that
  - The Input Weight We are Using to Drive the PMP CER is itself an Estimate Made by those who Put Together the CARD – When the Project is Completed, the Weight is Almost Sure to be Different, Probably Higher
  - The Percentage Factor we are Using to Estimate SE Cost is Only the Most Likely Factor, not Necessarily the Actual Factor for this Situation
- We Have to Consider the CER for SE Cost as

$$SE\ Cost = Factor \times PMP\ Cost,$$

where both “*Factor*” and “*PMP Cost*” have Probability Distributions

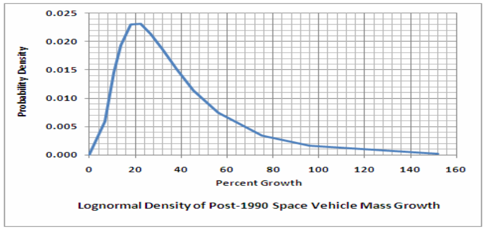
The Factor and PMP Cost Probability Distributions

- **We Model the Factor Distribution as the Triangular Distribution Derived Earlier from the Allard Study:**



- **Lacking More Specific Information, We Model Weight Growth as a Lognormal Distribution with Parameters**

P =	3.4621458	μ =	39.94
Q^2 =	0.4506876	σ =	30.14
Q =	0.6713327	Mode =	20.32



Derived from the Sugiyama-Yang Study Described on Charts 22-24

Establishing the Probability Distribution of SE Cost

- Note: It's Theoretically Possible that Weight "Growth" Could be Negative, but the Odds are Heavily Against it – We'll Ignore that Possibility in this Example
- We Generate a Random Sample of 10,000 Possible PMP Factor Values (F) from a Triangular Distribution with L = 31.0%, M = 45.5%, and H = 58.5%
- We Generate a Random Sample of 10,000 Possible Lognormally Distributed Weight-Growth Values (WG) with μ = 39.94 and σ = 30.14 (corresponding to normal distribution parameters P = 3.462 and Q = 0.671)
- Then, to Define the Probability Distribution of PMP Cost, we Calculate 10,000 Possible Values of

$$\text{PMP Cost} = F \times \{10.61+27.34\times[78\times(1+WG)]^{0.86}\}$$

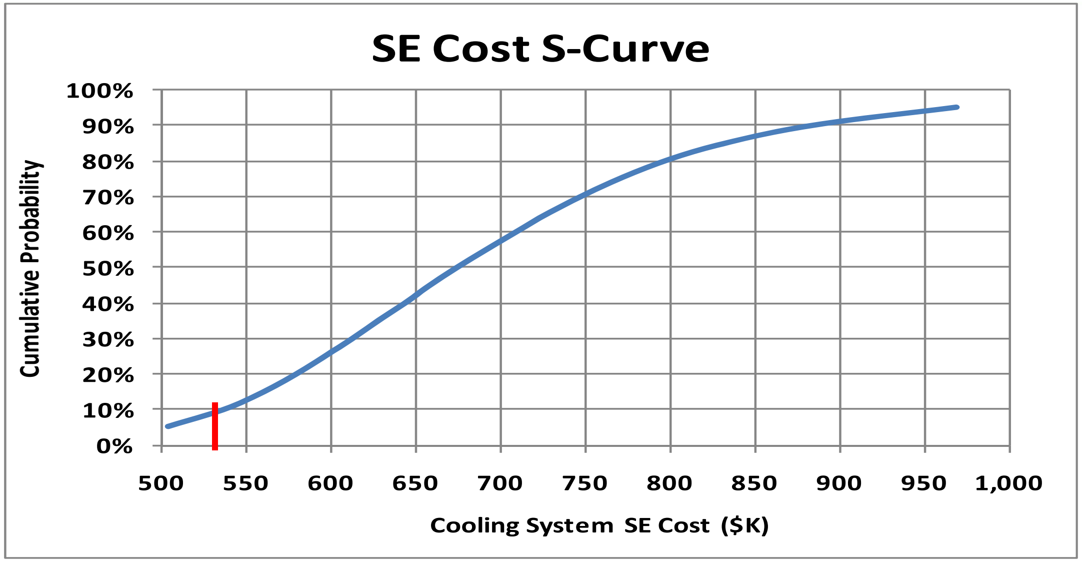
Simulation of Cooling System SE Cost Distribution

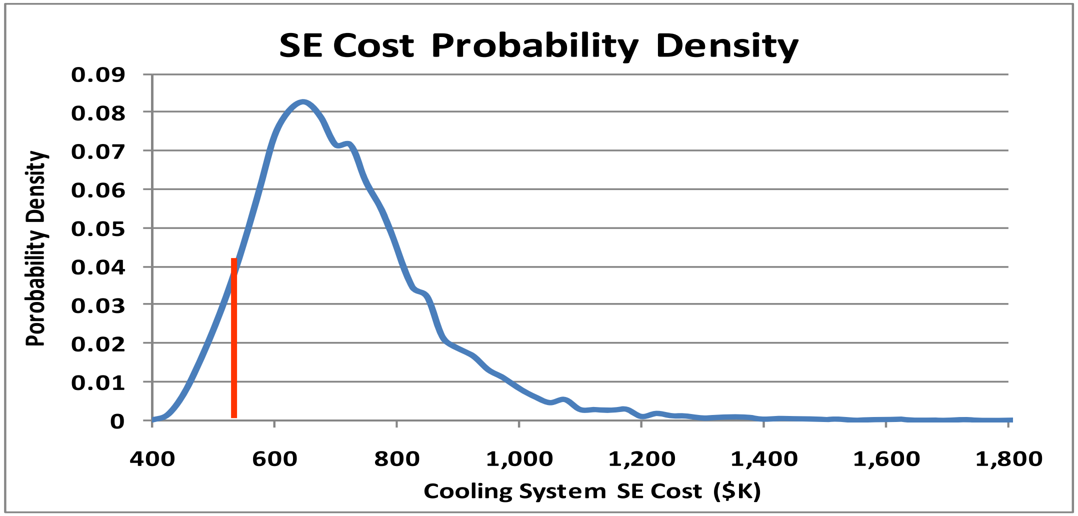
Random Number Trial	PMP Factor (F)	Weight (Pounds) with Growth $WWG = 78 \times (1+WG)$	CER-Based PMP Estimate ($K) $EST = 10.61 + 27.34 \times (WWG^{0.86})$	SE Cost ($K) Distribution $F \times EST$
1	0.38	103.48	1,488.22	567.79
2	0.46	85.47	1,264.27	581.95
3	0.39	98.41	1,425.77	560.86
4	0.35	130.14	1,810.33	640.21
5	0.41	132.01	1,832.48	755.09
6	0.32	102.71	1,478.78	479.73
7	0.44	86.03	1,271.22	557.32
8	0.45	145.45	1,990.96	886.07
•	•	•	•	•
•	•	•	•	•
•	•	•	•	•
9994	0.50	92.15	1,348.03	670.50
9995	0.44	91.49	1,339.82	595.75
9996	0.37	132.14	1,834.06	673.26
9997	0.52	95.84	1,394.00	730.70
9998	0.45	93.03	1,358.95	605.61
9999	0.52	90.47	1,327.04	696.23
10000	0.52	114.25	1,619.59	839.36

The Cooling System SE Cost Estimate and its Statistics

Mean =	$699.20K
Std Dev =	$153.59K

SE Cost ($K) Percentiles	
5th	503.99
9.06th (Naïve Estimate)	532.07
10th	537.98
15th	561.10
20th	580.30
25th	596.80
30th	613.26
35th	628.20
40th	644.16
45th	658.55
50th	674.88
60th	691.84
65th	709.47
70th	727.31
75th	747.84
80th	770.77
85th	797.22
90th	833.49
95th	884.89





Contents

- Current Status of Estimating Relationships
- Why Cost and Schedule Must be Modeled as Probability Distributions
- Why Cost and Schedule Drivers Must be Modeled as Probability Distributions
- What's Wrong with the Way Cost-on-Cost Estimating Relationships are Modeled and How to Fix It
- Applications
- Summary

Summary

- **Every Estimating Task Must be Treated as a Risk Analysis because …**
  - **Uncertainty Characterizes Almost All Cost and Schedule Drivers, as well as Estimates Based on Them**
  - **Arithmetic Roll-Up of "Point" Estimates in not Adequate – Cost Elements Must be Modeled Statistically and Summed by Monte-Carlo Sampling or Analytic Approximation)**
- **There are Two Basic Ways of Doing Cost**
  - **List All the Risk Issues Facing the Project and Assess their Combined Impact on Program Cost**
  - **Determine the Major Cost Drivers and Assess the Impact of Risks on Them (the drivers)**
- **Cost-Driver Risks Can be Modeled, for Example, as …**
  - **Weight, Lines-of-Code, etc., Growth when CER-Based Estimates are Calculated**
  - **Input Cost Uncertainty in the case of "Cost-on-Cost" CERs**
  - **Schedule Uncertainty when Schedule is the Cost Driver**

References

1. Allard, S., "Cost Factors for Cost Estimating and Analysis," Federal Aviation Administration, 2001.
2. Anderson, T., "Another Look at the Weight-Growth Study," The Aerospace Corporation, 2006.
3. Covert, R., "Errors-in-Variables Regression for CER Development," *Journal of Cost Analysis and Parametrics*, Spring 2008.
4. Covert, R., "Reducing Systemic Errors in Cost Models," ISPA/SCEA Joint International Conference, Noordwijk, Netherlands, 12-14 May 2008.
5. Gayek, J., L. Long, K. Bell, R. Hsu, and R. Larson, "Software Cost and Productivity Model," The Aerospace Corporation, Ground Systems Architecture Workshop, 2004.
6. Holchin, B., "Code Growth Methodology," 2006.
7. Jensen, R., "Estimating Software Growth," USAF Software Technology Support Center, Hill AFB UT, Space Systems Cost Analysis Group, 2008.
8. Jones, R. and P. Hardin, "Software Code Growth: A New Approach Based on Historical Analysis of Actuals," Navy/USMC Cost Analysis Symposium, 2007.
9. Reifer, D.J. "Software Cost, Quality, and Productivity Benchmarks," *The DoD Software Tech News* 7.2 (2004): 3-8, 17-19. http://www.compaid.com/caiinternet/ezine/Reifer-Benchmarks.pdf
10. Ross, M., "Software Size Uncertainty: The Effects of Growth and Estimation Variability," r2 Estimating, 2007
11. Sugiyama, R. and L. Yang, "Why Is It Difficult to Predict Mass Growth on Space Programs?," *Crosslink (The Aerospace Corp. Magazine)*, Fall 2006, http://www.aero.org/publications/crosslink/fall2006/06.html.

Acronyms

AFCAA	Air Force Cost Analysis Agency
ATP	Authority to Proceed (aka “Project Start”)
BTL	Below the Line
CARD	Cost Analysis Requirements Description
CER	Cost-Estimating Relationship
CoC	Cost on Cost
DSP	Defense Support Program
EIV	Errors in Variables
FTE	Full Time Equivalent
HR	Hour
HRS	Hours
LOE	Level of Effort
MPE/ZPB	Minimum Percentage Error / Zero Percentage Bias (= ZMPE)
PM	Program Management
PMP	Prime Mission Product
SBIRS	Space-Based Infrared System
SE	System Engineering
SER	Schedule-Estimating Relationship
SLOC	Source Lines of Code
SMC	(USAF) Space and Missile Systems Center
TRL	Technology-Readiness Level
USAF	United States Air Force
ZMPE	Zero Percentage Bias, Minimum Percentage Error (= MPE/ZPB)

Speaker Biographies

Dr. Stephen A. Book (sbook@mcri.com) recently vacated the position of Chief Technical Officer of MCR, LLC to concentrate on research and training. His responsibilities now include assisting MCR management with ensuring technical excellence of MCR products, services, and processes by encouraging process improvement, maintaining quality control, and training employees and customers in cost and schedule analysis and associated program-control disciplines. Dr. Book joined MCR in January 2001 after 21 years with The Aerospace Corporation, where he held the title "Distinguished Engineer" during 1996-2000 and served as Director, Resource and Requirements Analysis Department, during 1989-1995. At The Aerospace Corporation, he was a principal contributor to several Air Force cost studies of national significance, including the DSP/FEWS/BSTS/AWS/Brilliant Eyes Sensor Integration Study (1992) and the ALS/Spacelifter/EELV Launch Options Study (1993). Dr. Book is co-editor of the combined ISPA/SCEA technical journal, *The Journal of Cost Analysis and Parametrics*. He was the 2005 recipient of ISPA's Freiman Award for Lifetime Achievement and the 2010 recipient of SCEA's Award for Lifetime Achievement. Dr. Book earned his Ph.D. in mathematics, with concentration in probability and statistics, at the University of Oregon.

Amanda J. Feather (afeather@mcri.com) has a background in mechanical engineering and cost estimating for space programs. Currently located at MCR's El Segundo CA office, Ms. Feather supports the Air Force Cost Analysis Agency at the USAF Space and Missile Systems Center, Los Angeles, estimating costs and schedules for various space programs. Prior to joining MCR, she worked on national programs and power systems for military communication satellites at Boeing Satellite Systems. Ms. Feather earned her B.S. in Mechanical Engineering, with concentration in manufacturing, at Rensselaer Polytechnic Institute, New York.