Reducing Risk and Uncertainty in COSYSMO Size and Cost Drivers: Some Techniques for Enhancing Accuracy

Ricardo Valerdi
MIT; 77 Massachusetts Ave., Cambridge, MA 02139
rvalerdi@mit.edu
John E. Gaffney, Jr.
Lockheed Martin; One Curie Court, Rockville, MD 20850
j.gaffney@lmco.com

March 14, 2007
5th Conference on Systems Engineering Research

Overview

- This presentation describes how the COSYSMO systems engineering effort model/tool develops a probabilistic range of estimates rather than just a single point cost estimate.
  - This is the COSYSMO Risk or COSYSMO-R approach.
- The purpose of doing so is to provide systems engineers, managers, and other decision makers insight to make better decisions concerning such matters as the probability of a project's meeting its cost target.
- COSYSMO-R provides "risk" and "confidence" distributions for the labor and schedule or project duration estimates, based three-point values for each of its parameters that the user enters.
- Risk=Prob[actual value >target value]; the complementary cumulative distribution function (CCDF).

  Confidence=100%-Risk%=Prob[actual≥ target value]; the cumulative distribution function (CDF) of the cost.
- Note: these definitions apply to quantities for which "better" is smaller, e.g., effort/cost and project duration. They are reversed for cases in which "better" is larger, such as Mean-Time-Between Failure.

Basic or Academic COSYSMO Overview

- The academic COSYSMO model is implemented on an excel spreadsheet that provides an estimate of the total effort for five systems engineering activities over four life cycle phases
- The five activities* are:
  - Acquisition & Supply
  - Technical Management
  - System Design
  - Product Realization
  - Technical Evaluation
- The four phases** are:
  - Conceptualize
  - Develop
  - Operational Test & Evaluation
  - Transition to Operation

*from ANSI/EIA 632 **inspired by ISO 15288

Basic or Academic COSYSMO Overview, Contd.

- The fundamental equation implemented by COSYSMO (Valerdi 2005) and COSYSMO-R is:

  $PH = A*(SE)*\Pi Di$

  where:

  - PH=systems engineering person hours
  - A=unit effort constant
  - S=equivalent size, number of equivalent requirements
  - E=exponent
  - Di, i=1,2,….,14 are the cost driver values
- All of these parameters are considered to be mutually independent.

Probability Approximation Used In COSYSMO-R

- The COSYSMO-R risk assessment capability is implemented using three-point approximations; they are non-parametric, meaning that they are not derived as approximations to any particular distribution such as a Gamma or a Weibull.
- This in contrast to the use of Monte Carlo methods, in which a particular distribution is used and then a large number of instances are generated from it.
- COSYSMO-R does not generate such a large number of instances.
  - Rather, it generates an approximation to the distribution from the 3 point approximations to each variable. For example, if there are 4 (mutually independent) variables, the approximation has 81 values (=3x3x3x3).

Probability Approximation Used In COSYSMO-R, Contd.

- We used the approximation developed by Keefer and Bodily, the "extended Pearson-Tukey" method.
  - They evaluated 22 approximations, and found this one to be the best in terms of its ability to estimate the means and variances of various distributions.
- This method approximates a continuous distribution by a discrete one:

Fractile	Probability Assigned
0.05	0.185
0.50	0.630
0.95	0.185

The COSYMO-R User Enter Three-Point Estimates For:

- Model Parameters A and E
- Scope or Project Size Characteristics, Equivalent Size Drivers:
  - Number of System Requirements
  - Number of System Interfaces
  - Number of System-Specific Algorithms
  - Number of Operational Scenarios
- Cost/Performance Characteristics, Cost Drivers:
  - Requirements Understanding
  - Architecture Understanding
  - Level of Service Requirements
  - Migration Complexity
  - Technology Risk
  - Documentation
  - # and diversity of installations/platforms
  - # of recursive levels in the design
  - Stakeholder team cohesion
  - Personnel/team capability
  - Personnel experience/continuity
  - Process capability
  - Multi-site coordination
  - Tool Support

COSYSMO-R Example of Uncertainty Capture for Four Cost Drivers

- In this example, the range of values for each of four cost drivers is given in terms of Low, Likely , and High values that characterize the uncertainty in the estimator's belief in their values.

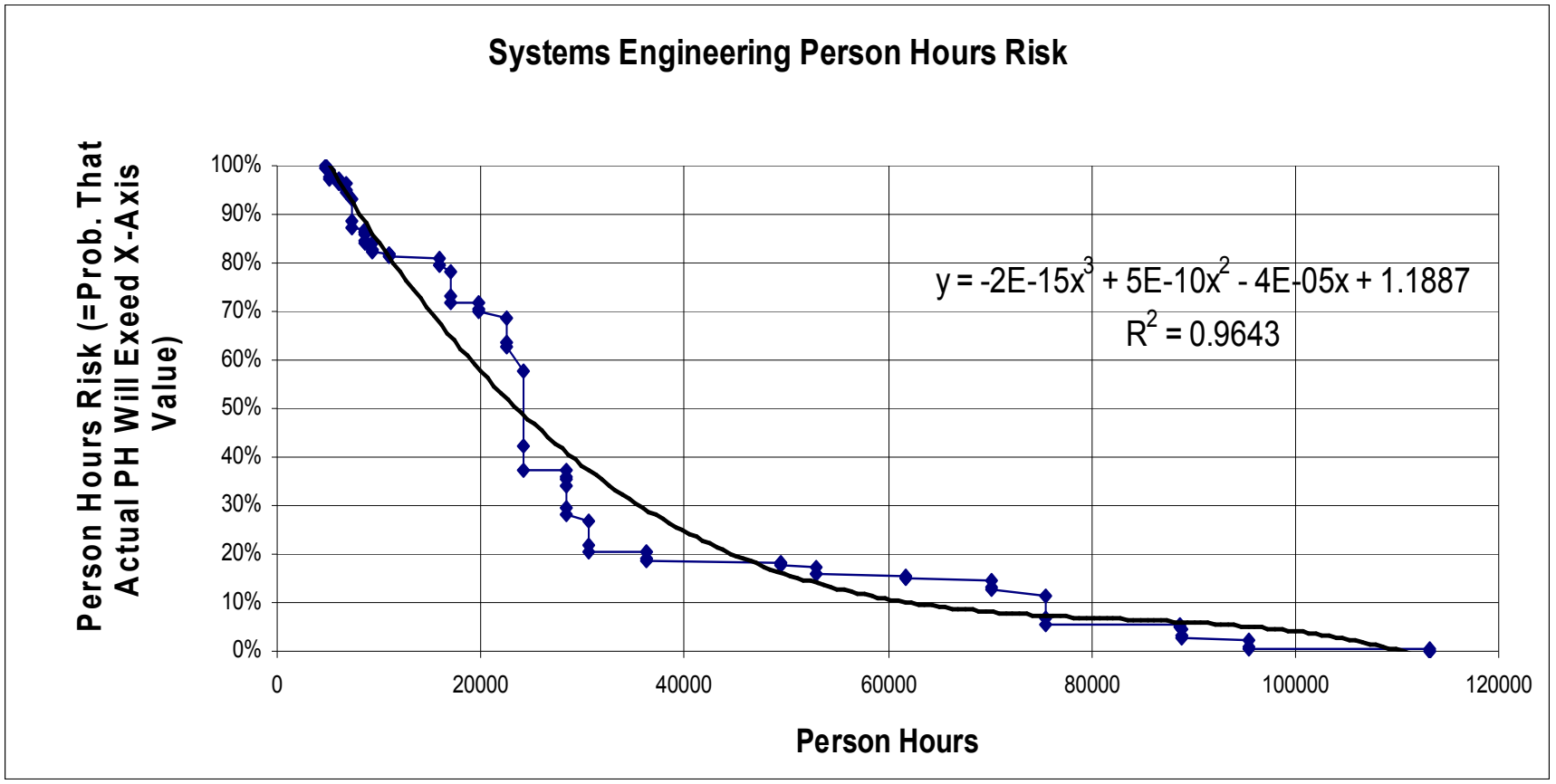
COSYSMO-R Example of Uncertainty Capture for Four Cost Drivers, Contd.

Driver Number	Driver Name	Values XL	VL	L	N	H	VH	XH
1	**Reqm'ts. Under.**	1.9	1.71	1.30	1.00	0.75	0.65	0.50
	Low Estimate			x				
	Likely Estimate				x			
	High Estimate						x	
2	**Technol. Risk**	0.5	0.68	1.26	1.00	1.3	1.75	2.00
	Low Estimate		x					
	Likely Estimate					x		
	High Estimate						x	
3	**Pers/Team Cap.**	1.59	1.5	1.12	1.00	0.87	0.68	0.62
	Low Estimate		x					
	Likely Estimate				x			
	High Estimate						x	
4	**Tool Support**	1.43	1.4	1.1	1.00	0.87	0.75	0.62
	Low Estimate				x			
	Likely Estimate					x		
	High Estimate						x	
						Most Likely Driver Product Value=		1.131

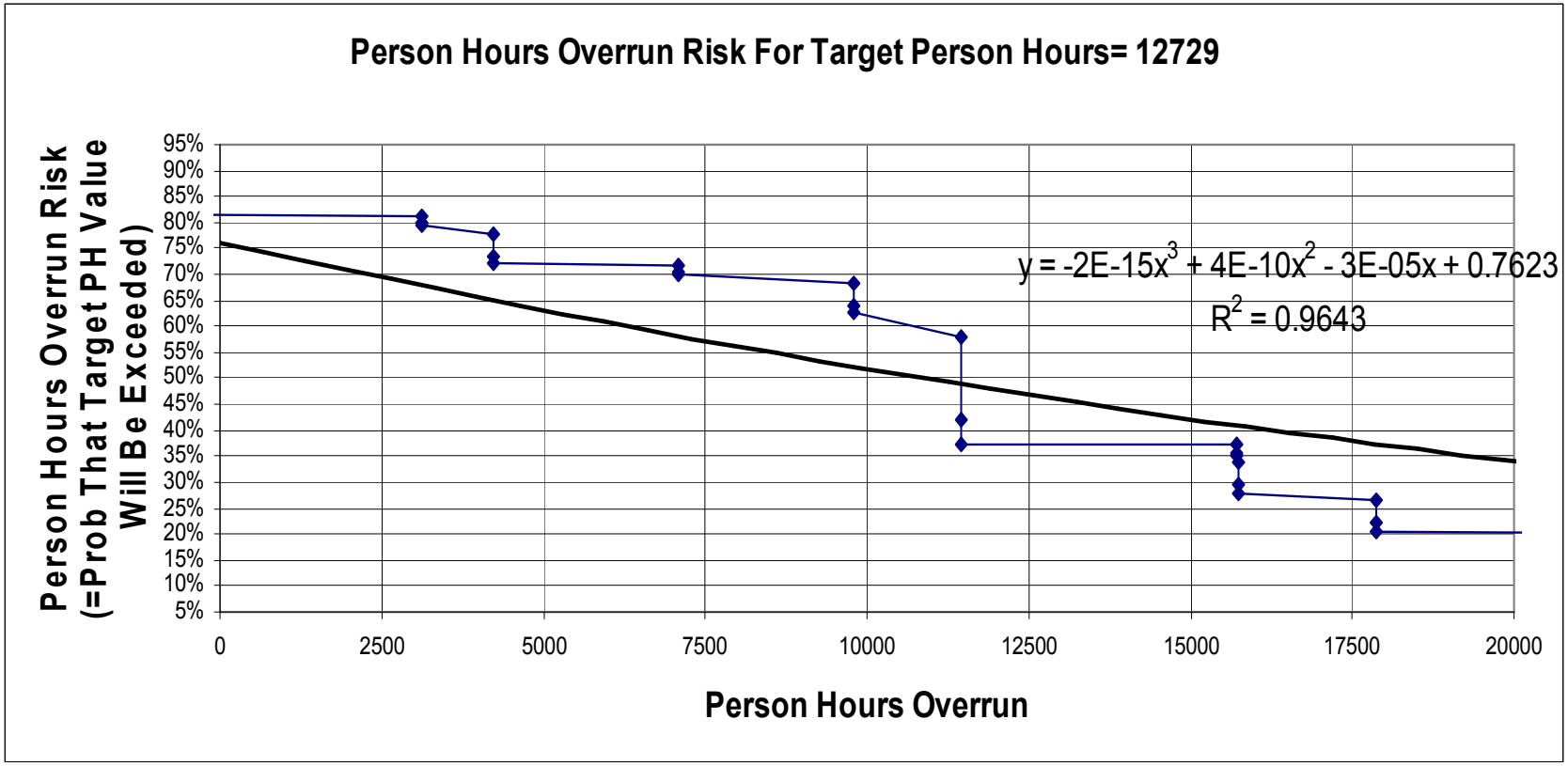
COSYSMO-R Example of Uncertainty Representation of Systems Engineering Person Hours Risk

- COSYMO-R generates a plot of the “risk” for systems engineering labor hours.
- Shown is both the discrete (point-by-point) representation as well as a smooth curve through it.
  - In some cases, the estimator might choose to show just the smooth curve to his audience.

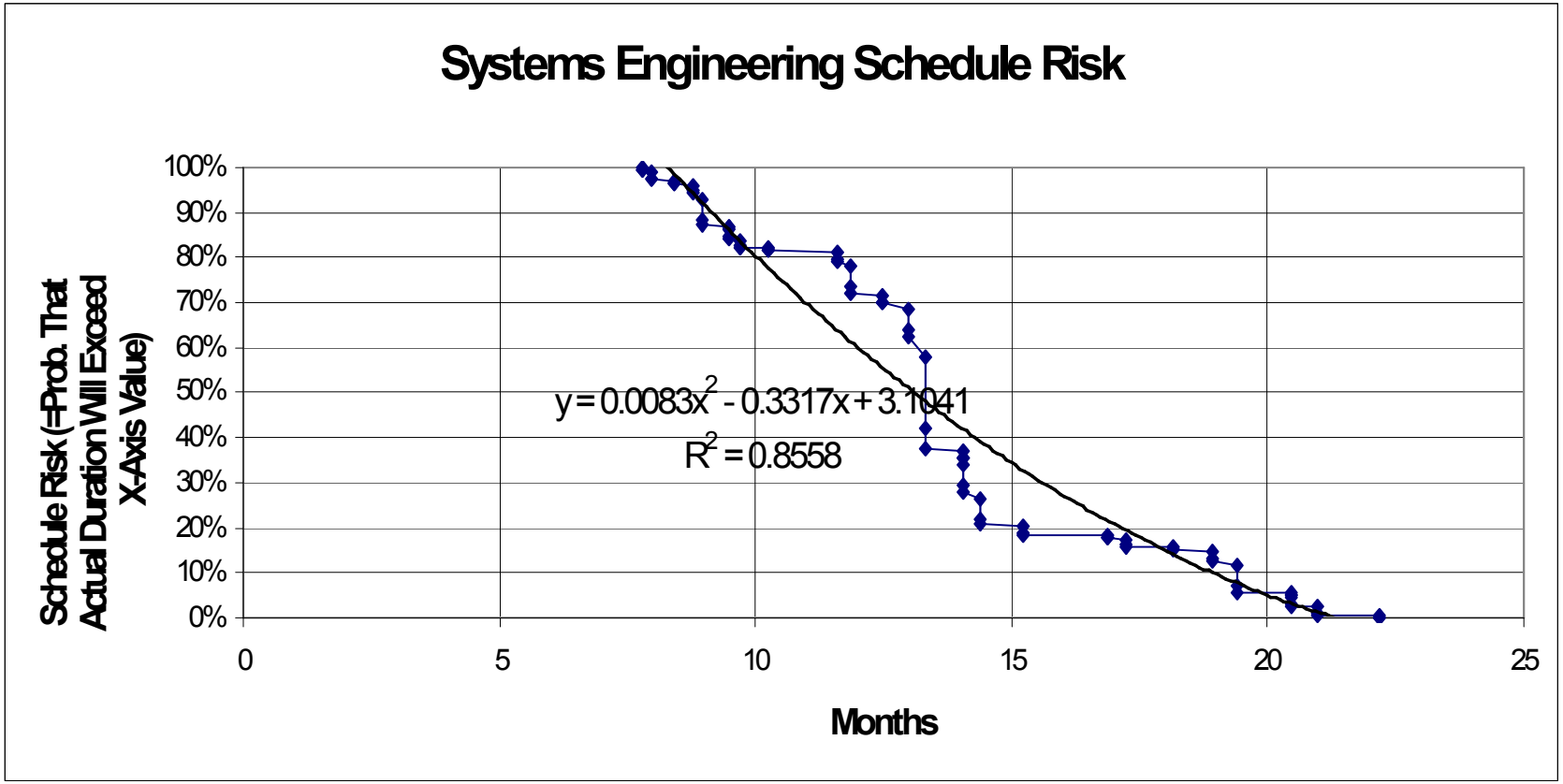
COSYSMO-R Example of Uncertainty (Risk) Representation of Systems Engineering Person Hours Risk



COSYSMO-R Example of Systems Engineering Person Hours Overrun Uncertainty (Risk) Representation



COSYSMO-R Example of Systems Schedule Uncertainty (Risk) Representation



Future Work

- There are other approaches to the estimation of risk in systems engineering.
  - COSYSMO can help to assess a risk profile for a project based on the combination of inputs.
- Pre-determined combinations of cost driver ratings can provide red flags for possible risks.
  - For example, if the architecture understanding cost driver is rated “Very Low” and the “Technology Maturity” is also rated “Very Low,” then this indicates a risk.
- Example: Ten high risk scenarios are provided in the figure on the next page.
  - They were obtained from a survey of twenty systems engineering experts who participated in a COSYSMO workshop. The numbers in each cell represent the number of votes received by that combination.
    - The two top risk areas were found to be requirements understanding and architecture understanding.

Future Work, Contd.

	SIZE	RQMT	ARCH	LSVC	MIGR	TRSK	DOCU	INST	RECU	TEAM	PCAP	PEXP	PROC	SITE	TOOL
SIZE (REQ + INTF + ALG + OPSC)		21	21	9	12	5	4	7	10	8	9	11	7	6	7
Requirements Understanding			17	9	7	8	3	5	9	5	10	8	5	4	1
Architecture Understanding				9	10	12	3	7	11	6	11	11	5	6	4
Level of Service Requirements (the ilities)					5	7	4	5	3	6	4	4	2	3	2
Migration Complexity (legacy system considerations)						8	1	10	1	4	7	7	3	5	4
Technology Risk (maturity of technology)							2	8	6	4	9	5	3	3	5
Documentation match to life cycle needs								2	3	4	4	2	6	2	3
Number and Diversity of Installations or Platforms									4	3	5	6	4	8	5
Number of Recursive Levels in the Design										4	8	7	7	2	5
Stakeholder Team Cohesion											7	9	3	8	3
Personnel/team capability												12	9	8	5
Personnel Experience and Continuity													10	8	3
Process Capability														5	8
Multisite Coordination															8
Tool Support															

high risk

medium risk

low risk

small x = 0.5; big X = 1

n = 19

Conclusions

- There is nothing as certain as uncertainty in cost estimation.
- The approach implemented in COSYSMO-R is useful for assessing the uncertainty in cost estimates and in quantifying the subjectivity involved in the estimation process.
- COSYSMO/COSYSMO-R can be used when doing “what if” scenarios in order to support the consideration of alternatives for project implementations.
- COSYSMO/COSYSMO-R can be used to help the enforcement of a systematic methodology in doing estimation. The estimator is forced to quantify his uncertainty in the values of key parameters and have them subjected to the scrutiny of others.