Joint Confidence Level of a Parametric Software Cost and Schedule Estimate:

Method and Example

Michael A. Ross
Technical Expert
Tecolote Research, Inc.
3601 Aviation Blvd., Suite 1600
Manhattan Beach, CA 90266-3756
310-536-0011
mross@tecolote.com
http://www.tecolote.com

[1,2]***Abstract***— This paper describes a *data-driven* approach to developing, describing, and calibrating a correlated software Cost Estimating Relationship (CER) and Schedule Estimating Relationship (SER) system of equations. It addresses CERs with correlated SERs that are probabilistic (i.e., provide a range of possible outcomes with associated probability of attainment). It then describes a simple method for determining joint and conditional probabilities (confidence levels) of estimates based on these relationships. The paper includes a practical example of developing a software CER and correlated SER from data collected by the Aerospace Corporation, then implementing the resulting model to examine several joint and conditional probability scenarios.

TABLE OF CONTENTS

1. INTRODUCTION

Background

Over 35 years of software development industry experience and project data tells us that cost and schedule estimating is essential to diligent affordability management, acquisition management, and project management. It also tells us that projects behave according to certain dynamic properties, that duration, effort, cost, and defects are all inexorably linked (correlated), that these correlations can be expressed as functions of product and project attributes, and that, prior to project

[2] 1st Edition, Version 05; May 10, 2011

completion, ***everything is uncertain*** (Ross, 2007a). Recognizing and managing the uncertainty (risk) goes a long way toward managing the project. Peter Bernstein (1996) writes

> "The revolutionary idea that defines the boundary between modern times and the past is the mastery of risk: the notion that the future is more than a whim of the gods and that men and women are not passive before nature. Until human beings discovered a way across that boundary, the future was a mirror of the past or the murky domain of oracles and soothsayers who held a monopoly over knowledge of anticipated events."

Dr. Paul Garvey, in his book *Probability Methods for Cost Uncertainty Analysis: A Systems Engineering Perspective* (2000), issued the following challenge to the cost analysis community:

> When cost uncertainty analyses are presented to decision-makers, questions often asked are ***"What is the chance the system can be delivered within cost and schedule?" "How likely might the point estimate cost be exceeded for a given schedule?" "How are cost reserve recommendations affected by schedule risk?"*** [author emphasis]
>
> During the past thirty years, techniques from univariate probability theory have been widely applied to provide insight into P(Cost ≤ x1) and P(Schedule ≤ x2).
>
> Although it has long been recognized that a system's cost and schedule are correlated, little has been applied from multivariate probability theory to study joint cost-schedule distributions. A multivariate probability model would provide analysts and decision-makers visibility into joint and conditional cost-schedule probabilities…

A primary goal of any thorough project estimating process should therefore be to not only yield estimated values for these metrics; it should also indicate whether or not these estimated values satisfy their corresponding goals (commitments) with some corresponding desired level of confidence (probability of success) (Ross, 2007a).

Recognition of this primary goal is exemplified in the 2004 Sambur-Teets memo, *"**High Confidence Estimates:** Estimate the software development and integration effort (staff hours), cost, and schedule at high (80-90%) confidence."* (Sambur, et al., 2004) and in the NASA Cost Estimating Handbook, *"As a general rule, cost estimates at NASA should be presented at the 70% confidence level. As an entire portfolio of Projects, the budget should be presented at the 80% confidence level"* (National Aeronautics and Space Administration (NASA), 2004). While many may take issue with prescribed confidence levels cited here; nonetheless, people are beginning to care about recognizing and managing uncertainty.

Recently the Space Systems Cost Analysis Group (SSCAG) Risk Subgroup, NASA, and the USAF Cost Analysis Agency (AFCAA) have been working separate efforts on joint cost and schedule risk analysis. One output of the SSCAG activity is an initial draft set of guidelines that includes the following (Druker, et al., 2009):

- ***Motivations –*** With the space systems community moving towards risk analysis methodologies that capture both cost and schedule metrics, it is important that common terminology and guidelines be developed to ensure a common frame of reference for discussions on this topic.
- ***Definition of Joint Cost & Schedule Risk Assessment –*** Joint Cost and Schedule Risk Assessment generates a joint bivariate probability distribution relating cost and schedule in a way that allows the analyst to determine the confidence level for meeting both target budgets and schedules simultaneously. This differs from traditional cost or schedule risk analysis in that program success is defined not as the probability of meeting one or the other, but as the probability of meeting both.
- ***Methods –*** Any method that ties cost and schedule risk together through the use of a bivariate distribution results in a joint cost and schedule risk assessment. The following are three examples of this type of analysis:
    - ***Parametric –*** Disjoint Cost and Schedule distributions are conflated into a bivariate distribution through the injection of correlation between the two
    - ***Buildup –*** An integrated master schedule is loaded with either costs or resources that translate to costs. Schedule uncertainty applied to tasks and discrete technical/program risks are added into the IMS. A Monte Carlo simulation is run on the schedule and from this a bivariate distribution of cost and schedule is produced
    - ***Estimate to Complete Projection –*** A cost distribution has a burn rate (with associated uncertainty) applied to it to determine an estimated finish date. A Monte Carlo simulation on the cost and burn rate distributions will result in a bivariate distribution of cost and schedule

Purpose and Scope

This paper describes a *data-driven* approach for developing, describing, and calibrating a correlated software Cost Estimating Relationship (CER) and Schedule Estimating Relationship (SER) system of equations. It addresses CERs with correlated SERs that are probabilistic (i.e., provide a range of possible outcomes with associated probability of attainment). It then describes a simple method for determining joint and conditional probabilities (confidence levels) of estimates based on these relationships. The paper includes a practical example of developing a software CER and correlated SER from data collected by the Aerospace Corporation, then implementing the resulting model to examine several joint and conditional probability scenarios.

2. Model Summary

Software CDER Regression Equations

These equations are used to support regression analysis of past project data (Ross, 2008).

$$\begin{aligned}
&CER\\
&\qquad \left[\mathbf{E} = b_1\mathbf{S}^{a_1}\right]_{\text{<dataset name>}} \rightarrow \left[\mathbf{E}\left(b_1\mathbf{S}^{a_1}\right)^{-1} = 1\right]_{\text{<dataset name>}} \qquad \textit{power regression}\\
&SER\\
&\qquad \left[\mathbf{T} = b_2\mathbf{E}^{a_2}\right]_{\text{<dataset name>}} \rightarrow \left[\mathbf{T}\left(b_2\mathbf{E}^{a_2}\right)^{-1} = 1\right]_{\text{<dataset name>}} \qquad \textit{power regression}\\
&CER \square SER \rightarrow CDER\\
&\qquad \left[\mathbf{E}\left(b_1\mathbf{S}^{a_1}\right)^{-1}\mathbf{T}\left(b_2\mathbf{E}^{a_2}\right)^{-1} = 1\right]_{\text{<dataset name>}}\\
&\textit{Letting} \quad \alpha_E \equiv \frac{1-a_2}{a_1} \quad \textit{and} \quad \alpha_T \equiv \frac{1}{a_1} \quad \textit{and} \quad \bar{D} \equiv \left(b_1 b_2\right)^{\frac{1}{a_1}} \quad \textit{yields}\\
&\qquad \left[\mathbf{E}^{\alpha_E}\mathbf{T}^{\alpha_T} = \bar{D}\mathbf{S}\right]_{\text{<dataset name>}} \qquad \textit{factor regression}
\end{aligned} \tag{1}$$

where

- $\mathbf{E} \equiv$ Vector of past project effort values from <**dataset name**>
- $\mathbf{T} \equiv$ Vector of past project duration values from <**dataset name**>
- $\mathbf{S} \equiv$ Vector of past project software size values from <**dataset name**>
- $\bar{D} \equiv$ Mean difficulty
- $\bar{I} \equiv$ Mean intensity
- $\alpha_E \equiv$ Effort exponent
- $\alpha_T \equiv$ Duration exponent
- $\gamma \equiv$ Economy exponent
- $b_1, b_2 \equiv$ Regression power function scale factor parameters
- $a_1, a_2 \equiv$ Regression power function exponent parameters

Random Variable forms of CDER Equations

These are a collection of probabilistic equations[3] that form the basis of a data-driven parametric joint cost and schedule estimating model (Ross, 2007a) (Valerdi, et al., 2009).

Bivariate Estimating Form

$$\left[\boldsymbol{E}^{\alpha_E}\boldsymbol{T}^{\alpha_T} = \boldsymbol{DS}\right]_{\text{<dataset name>}} \tag{2}$$

Work and Intensity Solved for Effort

$$\left[\boldsymbol{E} = \left(\boldsymbol{DST}^{-\alpha_T}\right)^{1/\alpha_E}\right]_{\text{<dataset name>}} \tag{3}$$

and

$$\left[\boldsymbol{E} = \boldsymbol{IT}^{\gamma}\right]_{\text{<dataset name>}} \tag{4}$$

[3] We use the ***Arial bold italic*** typeface to indicate a random variable.

Work and Intensity Solved for Duration

$$\left[\boldsymbol{T} = \left(\boldsymbol{DSE}^{-\alpha_E}\right)^{1/\alpha_T}\right]_{\text{<data set name>}} \tag{5}$$

and

$$\left[\boldsymbol{T} = \left(\frac{\boldsymbol{E}}{\boldsymbol{I}}\right)^{\frac{1}{\gamma}}\right]_{\text{<data set name>}} \tag{6}$$

Work and Intensity Calibration

$$\left[\boldsymbol{D} = \frac{\boldsymbol{E}^{\alpha_E}\boldsymbol{T}^{\alpha_T}}{\boldsymbol{S}}\right]_{\text{<data set name>}} \tag{7}$$

and

$$\left[\boldsymbol{I} = \frac{\boldsymbol{E}}{\boldsymbol{T}^{\gamma}}\right]_{\text{<data set name>}} \tag{8}$$

Intensity-Correlated CER and SER Equations

These intensity-correlated CER and SER equations facilitate finding the appropriate distributions of effort and duration at some known intensity.

Cost (Effort) Estimating Relationship (CER)

$$\left[\boldsymbol{E} = \left(\boldsymbol{I}^{\alpha_T}\left(\boldsymbol{DS}\right)^{\gamma}\right)^{1/(\gamma\alpha_E+\alpha_T)}\right]_{\text{<data set name>}} \tag{9}$$

Schedule (Duration) Estimating Relationship (SER)

$$\left[\boldsymbol{T} = \left(\boldsymbol{I}^{-\alpha_E}\boldsymbol{DS}\right)^{1/(\gamma\alpha_E+\alpha_T)}\right]_{\text{<data set name>}} \tag{10}$$

Intensity-Correlated CER and SER Equations for Intensity

We provide solved-for-intensity forms of the CER and SER equations to facilitate finding intensity associated with some particular solution or distribution of solutions. These equations make it possible to find the duration that corresponds (correlates) to a particular effort solution and vice versa.

Cost (Effort) Estimating Relationship (CER)

$$\left[\boldsymbol{I} = \left(\boldsymbol{E}^{\gamma\alpha_E+\alpha_T}\left(\boldsymbol{DS}\right)^{-\gamma}\right)^{1/\alpha_t}\right]_{\text{<data set name>}} \tag{11}$$

Schedule (Duration) Estimating Relationship (SER)

$$\left[\boldsymbol{I} = \left(\boldsymbol{T}^{\gamma\alpha_E+\alpha_T}\left(\boldsymbol{DS}\right)^{-1}\right)^{-1/\alpha_E}\right]_{\text{<data set name>}} \tag{12}$$

Intensity as a Function of Some Percentage of Mean Duration

This equation finds the value of intensity I that corresponds to an effort-duration solution pair where the duration is some given percentage of the mean data set duration $\bar{T}$.

$$\%\bar{T} = \left(\frac{\bar{I}}{\bar{I}_{\%\bar{T}}} \right)^{\frac{\alpha_E}{\gamma\alpha_E + \alpha_T}}$$
$$\therefore I_{\%\bar{T}} = \bar{I} \left(\frac{1}{\%\bar{T}} \right)^{\frac{\gamma\alpha_E + \alpha_T}{\alpha_E}} \quad \textbf{(13)}$$

The remainder of this paper describes the basis of these equations and shows examples of their application.

3. Developing Software CERs with Correlated SERs

Independent Cost Estimating Relationship (CER) Form

We first propose a software Cost Estimating Relationship (CER) of the power form $y = bx^a$

$$E = b_1 S^{a_1} \quad \textbf{(14)}$$

where

$E \equiv$ total effort or labor (person-months or person-hours)

$S \equiv$ total software size (Effective Source Lines of Code or ESLOC)

Initial CER Regression

Given a list of historical project size values $\mathbf{S}$ and a corresponding list of historical effort values $\mathbf{E}$, we choose a power estimating relationship of the form $y = bx^a$ using *zero-percent-bias minimum-percent-error* (ZPB-MPE) general regression implemented in Microsoft Excel using the Solver iterative solution add-in. This regression can be stated as the transform

$$\left[\mathbf{Y}, \mathbf{X} \rightarrow b, a \right]_{\textbf{<data set name>}} \quad \textbf{(15)}$$

where

$$\mathbf{Y} \leftarrow \mathbf{E}, \mathbf{X} \leftarrow \mathbf{S}, b \rightarrow b_1, a \rightarrow a_1$$

The percentage bias $\%BIAS$ of an estimating relationship $y = \boldsymbol{f}(x)$ is generally defined as the mean (arithmetic average) of the percentage differences between the actual outcome values and their corresponding estimated values.

$$\%BIAS \equiv \frac{1}{N} \sum_{i=1}^{N} \left[\frac{Y_i - \boldsymbol{f}(X_i)}{\boldsymbol{f}(X_i)} \right] \quad \textbf{(16)}$$

To address the *zero-percent-bias* constraint that is part of the ZPB-MPE regression strategy, we explicitly constrain the coefficient b by the function of a that assumes the estimating relationship's percentage bias $\%BIAS$ be equal to zero.

$$\begin{aligned}
&\%BIAS(a \mid b) = 0 = \frac{1}{N}\sum_{i=1}^{N}\frac{Y_i - \boldsymbol{f}(X_i)}{\boldsymbol{f}(X_i)} = \sum_{i=1}^{N}\frac{Y_i - bX_i^{\,a}}{bX_i^{\,a}} \\
&0 = \sum_{i=1}^{N}\left(\frac{Y_i}{bX_i^{\,a}} - \frac{bX_i^{\,a}}{bX_i^{\,a}}\right) = \sum_{i=1}^{N}\frac{Y_i}{bX_i^{\,a}} - \sum_{i=1}^{N}\frac{bX_i^{\,a}}{bX_i^{\,a}} = \frac{1}{b}\sum_{i=1}^{N}\frac{Y_i}{X_i^{\,a}} - \sum_{i=1}^{N}1 = \frac{1}{b}\sum_{i=1}^{N}\frac{Y_i}{X_i^{\,a}} - N \\
&N = \frac{1}{b}\sum_{i=1}^{N}\frac{Y_i}{X_i^{\,a}} \\
&\therefore b = \frac{1}{N}\sum_{i=1}^{N}\frac{Y_i}{X_i^{\,a}}
\end{aligned} \quad \textbf{(17)}$$

The percentage Standard Error of Estimate $\%SEE$ is generally defined as the root mean squared (RMS) percentage difference between the actual outcome values and their corresponding estimated values

$$\%SEE = \sqrt{\frac{1}{N-F}\sum_{i=1}^{N}\left(\frac{y_i - \boldsymbol{f}(X_i)}{\boldsymbol{f}(X_i)}\right)^2} \quad \textbf{(18)}$$

where F specifies the number of parameters in the particular estimating relationship (degrees of freedom).

To address the *minimum-percent-error* constraint that is part of the ZPB-MPE regression strategy, we set up Microsoft Excel Solver such that the objective (target) is to minimize $\%SEE$ (instantiated with our chosen estimating relationship form $y = bx^a$) by elaborating the parameter b with our *zero-percent-bias* constraint in Equation (17) and by varying a (initialized in Solver to a value of one).

$$\begin{aligned}
&\boldsymbol{\%SEE} = \sqrt{\frac{1}{N-F}\sum_{i=1}^{N}\left[\left(\frac{Y_i - \boldsymbol{f}(X_i)}{\boldsymbol{f}(X_i)}\right)^2\right]} = \sqrt{\frac{1}{N-2}\sum_{i=1}^{N}\left[\left(\frac{Y_i - bX_i^{\,a}}{bX_i^{\,a}}\right)^2\right]} \\
&\therefore \boldsymbol{\%SEE} = \sqrt{\frac{1}{N-2}\sum_{i=1}^{N}\left[\left(\frac{Y_i - \left(\frac{1}{N}\sum_{i=1}^{N}\left[\frac{Y_i}{X_i^{\,a}}\right]\right)X_i^{\,a}}{\left(\frac{1}{N}\sum_{i=1}^{N}\left[\frac{Y_i}{X_i^{\,a}}\right]\right)X_i^{\,a}}\right)^2\right]}
\end{aligned} \quad \textbf{(19)}$$

When Microsoft Solver is finished, it has found the value of a that yields the smallest value for $\%SEE$. The corresponding value for b is simply found by solving Equation (17) using the final value found for a.

Independent Schedule Estimating Relationship (SER) Form

We next propose a software Schedule Estimating Relationship (CER) of the form

$$T = b_2 E^{a_2} \tag{20}$$

where

$T \equiv$ total duration or schedule (elapsed calendar months)

$E \equiv$ total effort or labor (person-months or person-hours)

Note that ***the SER is not a function of size*** but rather of effort since we are ultimately interested in correlating effort and time, not just deriving two independent functions of size.

Initial SER Regression

Given a list of historical project effort values $\mathbf{E}$ and a corresponding list of historical duration values $\mathbf{T}$, we choose a power estimating relationship of the form $y = bx^a$ using ZPB-MPE general regression implemented in Microsoft Excel using the Solver iterative solution add-in. This regression can be stated as the transform

$$\left[\mathbf{Y}, \mathbf{X} \to b, a\right]_{\langle\mathbf{dataset\,name}\rangle} \tag{21}$$

where

$\mathbf{Y} \leftarrow \mathbf{T}, \mathbf{X} \leftarrow \mathbf{E}, b \to b_2, a \to a_2$

and can be performed using the same method as described above for the initial CER regression.

Correlating a CER with its Corresponding SER

We now have a system of two nonlinear simultaneous equations: the CER $E = b_1 S^{a_1}$ and its corresponding SER $T = b_2 E^{a_2}$. Logically conflating this system of nonlinear simultaneous equations yields a single equation that is true if and only if both the CER and the SER are true. We hereinafter refer to this new relationship as a ***Cost and Duration Relationship (CDER)***.

$$\begin{aligned}
&E = b_1 S^{a_1} \quad \text{and} \quad T = b_2 E^{a_2} \\
&\left(ES^{-a_1} b_1^{-1} = 1\right) \wedge \left(TE^{-a_2} b_2^{-1} = 1\right) \\
&ES^{-a_1} b_1^{-1} TE^{-a_2} b_2^{-1} = 1 \\
&\therefore E^{\frac{1-a_2}{a_1}} T^{\frac{1}{a_1}} = \left(b_1 b_2\right)^{\frac{1}{a_1}} S
\end{aligned} \tag{22}$$

Consolidating the CDER Coefficients and Exponents

In order to visually simplify Equation (22) we define the following new variables

$$\bar{D}_{initial} \equiv \left(b_1 b_2\right)^{\frac{1}{a_1}} \tag{23}$$

$$\alpha_{E\_initial} \equiv \frac{1 - a_2}{a_1} \tag{24}$$

$$\alpha_{T\_initial} \equiv \frac{1}{a_1} \quad \textbf{(25)}$$

$$\bar{I} \equiv \left(\frac{1}{b_2}\right)^{\frac{1}{a_2}} \quad \textbf{(26)}$$

$$\gamma \equiv \frac{1}{a_2} \quad \textbf{(27)}$$

where

$\bar{D}_{initial} \equiv$ initial (first pass) mean difficulty

$\alpha_{E\_initial} \equiv$ effort exponent

$\alpha_{T\_initial} \equiv$ duration exponent

$\bar{I} \equiv$ mean intensity

$\gamma \equiv$ economy exponent

Substituting the new variables defined in Equations (23), (24), and (25) for their equivalents in Equation (22) yields

$$E^{\alpha_{E\_initial}} T^{\alpha_{T\_initial}} = \bar{D}_{initial} S \quad \textbf{(28)}$$

Solving Equation (24) for a_2 yields

$$a_2 = 1 - a_1 \alpha_{E\_initial} \quad \textbf{(29)}$$

Solving Equation (25) for a_1 yields

$$a_1 \equiv \frac{1}{\alpha_{T\_initial}} \quad \textbf{(30)}$$

Substituting the equivalent of a_1 in Equation (30) for a_1 Equation (29) yields

$$a_2 = 1 - \frac{1}{\alpha_{T\_initial}} \alpha_{E\_initial} \quad \therefore \quad a_2 = \frac{\alpha_{T\_initial} - \alpha_{E\_initial}}{\alpha_{T\_initial}} \quad \textbf{(31)}$$

Substituting the equivalent of a_2 in Equation (31) for a_2 in Equation (27) yields

$$\gamma = \frac{1}{\left(\frac{\alpha_{T\_initial} - \alpha_{E\_initial}}{\alpha_{T\_initial}}\right)} \quad \therefore \quad \gamma = \frac{\alpha_{T\_initial}}{\alpha_{T\_initial} - \alpha_{E\_initial}} \quad \textbf{(32)}$$

Solving Equation (20) for effort yields

$$E = \left(\frac{1}{b_2}\right)^{\frac{1}{a_2}} T^{\frac{1}{a_2}} \quad \textbf{(33)}$$

Substituting the equivalent of $\left(\frac{1}{b_2}\right)^{\frac{1}{a_2}}$ in Equation (26) and the equivalent of $\frac{1}{a_2}$ in Equation (27) for $\left(\frac{1}{b_2}\right)^{\frac{1}{a_2}}$ and $\frac{1}{a_2}$ in into Equation (33) yields

$$E = \bar{I}T^{\gamma} \quad \text{or} \quad T = \left(\frac{E}{\bar{I}}\right)^{\frac{1}{\gamma}} \quad \text{or} \quad \bar{I} = \frac{E}{T^{\gamma}} \tag{34}$$

Improving the Regressed CDER Exponent Values

Equation (28) represents an initial cut at a CDER for software development, the result of multiplicatively combining a CER and an SER. We can improve upon the initial exponent values $\alpha_{E\_initial}$ and $\alpha_{T\_initial}$ resulting from the initial CER and SER regressions by performing one final CDER regression that yields final exponent values α_E and α_T and a final value for mean difficulty $\bar{D}$. We can show how γ correlates the exponent values $\alpha_{E\_initial}$ and $\alpha_{T\_initial}$ by solving Equation (32) for $\alpha_{T\_initial}$.

$$\alpha_{T\_initial} = \frac{\gamma}{\gamma - 1}\alpha_{E\_initial} \tag{35}$$

We wish to maintain this correlation during the final regression; therefore we impose the constraint

$$\alpha_T = \frac{\gamma}{\gamma - 1}\alpha_E \tag{36}$$

In order to avoid the necessity of varying both α_E and α_T during the final regression and thus risking Microsoft Excel Solver finding a non-optimal (local minima) solution, we choose to vary only exponent α_E and constrain exponent α_T by Equation (36).

Final Regression

Given a list of historical project size values $\mathbf{S}$, a list of corresponding historical effort values $\mathbf{E}$, a list of corresponding historical duration values $\mathbf{T}$, and initial exponent values $\alpha_{E\_initial}$ and $\alpha_{T\_initial}$, we choose a factor estimating relationship of the form $y = bx$ using ZPB-MPE general regression implemented in Microsoft Excel using the Solver iterative solution add-in. This regression can be stated as the transform

$$\left[\mathbf{Y}, \mathbf{X} \rightarrow b\right]_{\text{<dataset name>}} \tag{37}$$

where

$$\mathbf{Y} \equiv \mathbf{E}^{\alpha_E}\mathbf{T}^{\alpha_t}, \mathbf{X} \equiv \mathbf{S}, \bar{D} \equiv b$$

The goal of this regression is to instantiate the general form Cost and Duration Estimating Relationship (CDER)

$$E^{\alpha_E} T^{\alpha_T} = \bar{D} S \qquad \textbf{(38)}$$

with the given historical data.

The coefficient b in this regression can be explicitly constrained by the assumption that the CDER's percentage bias $\%BIAS$ be equal to zero.

$$\boldsymbol{\%BIAS}(b) = 0 = \sum_{i=1}^{N}\left[\frac{y_i - bx_i}{bx_i}\right]$$

$$0 = \sum_{i=1}^{N}\left[\frac{y_i}{bx_i} - \frac{bx_i}{bx_i}\right] = \sum_{i=1}^{N}\left[\frac{y_i}{bx_i}\right] - \sum_{i=1}^{N}\left[\frac{bx_i}{bx_i}\right] = \frac{1}{b}\sum_{i=1}^{N}\left[\frac{y_i}{x_i}\right] - \sum_{i=1}^{N}[1] = \frac{1}{b}\sum_{i=1}^{N}\left[\frac{y_i}{x_i}\right] - N \qquad \textbf{(39)}$$

$$N = \frac{1}{b}\sum_{i=1}^{N}\left[\frac{y_i}{x_i}\right]$$

$$\therefore b = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{y_i}{x_i}\right]$$

We can now solve for the exponent α_E by setting up Microsoft Excel Solver such that the objective (target) is to minimize $\%SEE$ (the root mean squared (RMS) percentage difference between the actual outcome and the estimate) held to the zero-bias-constraint explicit solution for b.

$$\%SEE = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left[\left(\frac{y_i - bx_i}{bx_i}\right)^2\right]} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left[\left(\frac{y_i - \left(\frac{1}{N}\sum_{i=1}^{N}\frac{y_i}{x_i}\right)x_i}{\left(\frac{1}{N}\sum_{i=1}^{N}\frac{y_i}{x_i}\right)x_i}\right)^2\right]} \qquad \textbf{(40)}$$

and varying α_E (initialized to $\alpha_{E\_initial}$) while replacing (constraining) α_T with its equivalent in Equation (36).

The results of this final regression are optimized values for the exponent α_E and the mean difficulty $\bar{D}$. The corresponding optimized value for the exponent α_T can be found by solving Equation (36) using the optimized value for α_E (note that the value of the exponent γ is already known from the SER regression and Equation (21). The corresponding difficulty list $\mathbf{D}$ can now be populated using the list form of Equation (38) solved for $\mathbf{D}$.

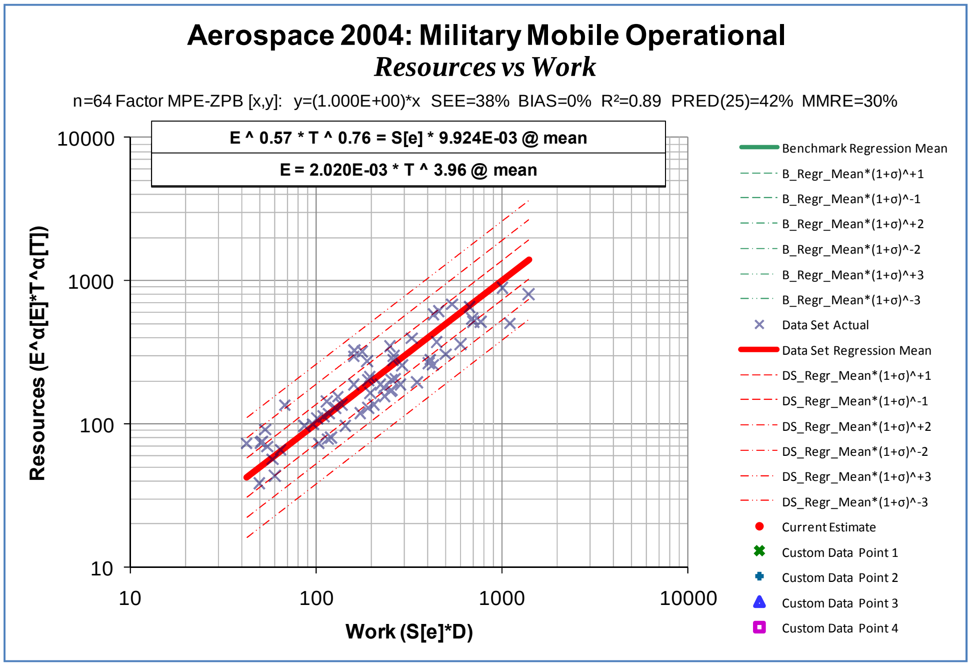
$$\mathbf{D} = \frac{\mathbf{E}^{\alpha_E}\mathbf{T}^{\alpha_T}}{\mathbf{S}} \qquad \textbf{(41)}$$

and the corresponding intensity list $\mathbf{I}$ can now be calculated using the list form of Equation (34) solved for $\mathbf{I}$.

$$\mathbf{I} = \frac{\mathbf{E}}{\mathbf{T}^{\gamma}} \qquad \textbf{(42)}$$

Performing the above-described regression process on an example set of stratified past project data (Aerospace Corporation, 2004) yields the results shown in Figure 1 below.

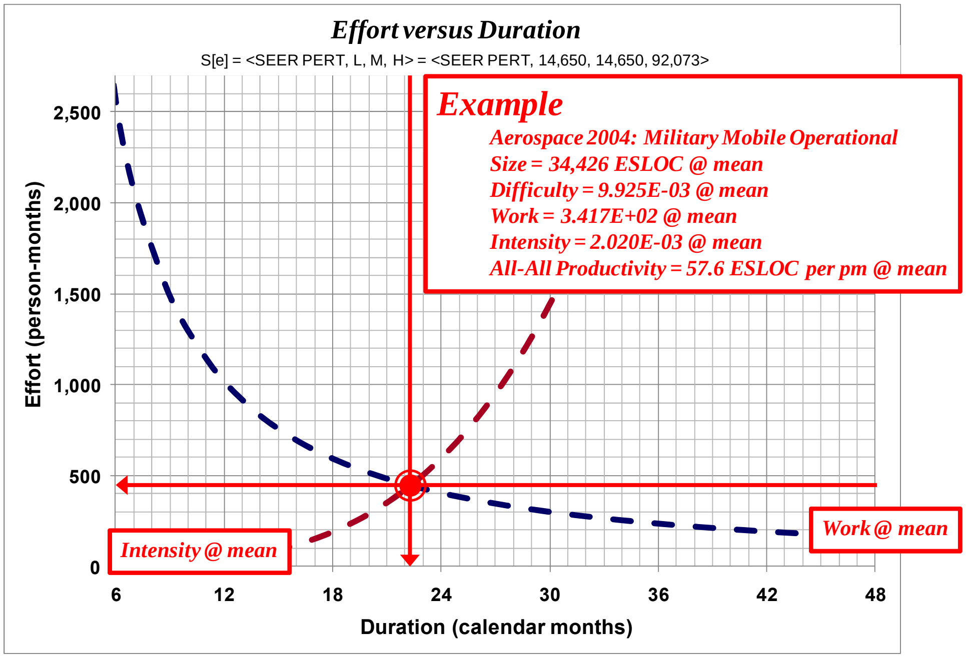
Figure 1: *Example CDER Regression Results*



If we elaborate Equation (38) and the solved-for-effort form of Equation (34) with the example regression values above, provide a distribution for size[4], and graph the results, we get Figure 2 below.

[4] Size in this example is given to be a contractor point estimate with 20% estimate maturity of 25,000 New DSLOC, growth-adjusted according to (Ross, 2011a), which yields a SEER PERT distribution parameter triple of <Least,Likely,Most> = <14650,14650,92073> with a distribution mean value of 34,426 DSLOC. In this example we make the simplifying assumption that all software to be developed is New DSLOC in order to avoid rework considerations; therefore, ESLOC equals DSLOC.

Figure 2: *Work Function (CDER) at a Particular Intensity*



4. Incorporating Probability in CDER Mathematics

Uncertainty about Size, Difficulty, Effort, and Duration

Up until this point, we have been treating the independent variables software content (size) S, difficulty D, and the size*difficulty product which we will call the ***work*** Ψ as either certain; i.e., single-point values or as lists of samples. Unfortunately, until project completion, we have exact values for neither the size nor the difficulty of the project being estimated; these values are ***uncertain***; i.e., they have a range of possible outcomes. We, therefore, choose to represent size and difficulty as random variables $\boldsymbol{S}$ and $\boldsymbol{D}$, and elaboration becomes modeling the distribution function for each.[5] From this we define the size*difficulty product (the work distribution $\boldsymbol{\Psi}$) as

$$\boldsymbol{\Psi} \equiv \boldsymbol{SD} \tag{43}$$

The choice of specific distributions for $\boldsymbol{S}$ and for $\boldsymbol{D}$ is a subject worthy of debate and a future paper. The author currently takes the position that when the data from a statistically-significant number of past projects exists, it is best to create a discrete mapping between the metric's range values and their corresponding cumulative probabilities[6] (i.e., a Cumulative Distribution Function (CDF) array) rather than assume some mathematically defined distribution. For the purposes of this paper, difficulty $\boldsymbol{D}$ is modeled this way, the range values coming directly from past project data; however, size $\boldsymbol{S}$ is modeled as a SEER PERT distribution (Ross, 2011a). Regardless of the distributions chosen, we need some way to determine the CDF $F_{\boldsymbol{\Psi}}(\psi)$ of the size*difficulty product $\boldsymbol{\Psi}$. Finding a neat closed-form CDF that is the convolved product of two

[5] We use the Arial bold italic typeface to indicate a random variable.

[6] Tecolote's ACEIT tool refers to this as a *Custom CDF*.

random variables, each distribution of which is described as some mathematical transform, is problematic at best. On the other hand, if the random variables are described as Cumulative Distribution Function (CDF) arrays, developed either from historical sample data or from a selected distribution's mathematics, then Monte Carlo methods can be used to determine the CDF array of the convolved product. This process is summarized as follows:

(1) Create ***randomly-ordered*** M -element lists $\mathbf{S}$ and $\mathbf{D}$ of distributed possible outcomes for each of size $\boldsymbol{S}$ and difficulty $\boldsymbol{D}$. For $\boldsymbol{S}$ we simply use SEER PERT mathematics (Ross, 2011a) to generate an M -element CDF array and then *shuffle* the range values S_i (see paragraph c below). For random variables such as $\boldsymbol{D}$ that are modeled directly from historical data as CDF arrays we must:

a. Perform a rank and percentile process on the N -element list of past project difficulty values $\mathbf{D}$ to yield a CDF array (N rows, 2 columns) where each row contains a range value D_i in the first column and its corresponding percentile value p_i in the second column.

$$\mathbf{D}_{\mathbf{CDF}} = \begin{bmatrix} D_1 & p_1 \\ D_2 & p_2 \\ \vdots & \vdots \\ D_N & p_N \end{bmatrix} \qquad \textbf{(44)}$$

b. Create a new expanded M -element CDF array $\mathbf{D}'_{\mathbf{CDF}}$ by interpolating M ascending-sorted uniformly distributed probability values $p \in [0,1]$ into the original past project CDF array. This produces an M -element CDF array that exactly models the past project data. The examples in this paper assume $M = 2001$ (the number of Monte Carlo *draws* used to model distributions for the paper's examples).

$$\mathbf{D}'_{\mathbf{CDF}} = \begin{bmatrix} D_1 & p_1 \\ D_2 & p_2 \\ \vdots & \vdots \\ D_M & p_M \end{bmatrix} \qquad \textbf{(45)}$$

c. Shuffle the M -element CDF vector. This can be done by adding a column of random numbers to the CDF array $\mathbf{D}'_{\mathbf{CDF}}$ and then sorting the rows of the new array $\mathbf{D}''_{\mathbf{CDF}}$ using the column of random numbers as the sort key. Note that the range values column of the CDF array $\mathbf{D}''_{\mathbf{CDF}}$ is now randomly ordered.

(2) Compute an M -element list $\mathbf{\Psi}$ for the size*difficulty product (work) $\boldsymbol{\Psi}$ that is the result of scaling a randomly-ordered list of difficulty values $\mathbf{D}$ by a randomly-ordered list of size values $\mathbf{S}$ such that

$$\mathbf{\Psi} = \mathbf{SD} \equiv \begin{bmatrix} S_1 D_1 \\ S_2 D_2 \\ \vdots \\ S_M D_M \end{bmatrix} \quad \textbf{(46)}$$

(3) Sort the work list $\mathbf{\Psi}$ in ascending order to get $\mathbf{\Psi}'$.

(4) Add an M -element column to $\mathbf{\Psi}'$ that contains the quantile of each corresponding range value ψ_i in the sorted work list $\mathbf{\Psi}$[7]; the result is the CDF array $\mathbf{\Psi}_{\mathbf{CDF}}$ of the work random variable $\boldsymbol{\Psi}$ (note that each column of the CDF array is in ascending order).

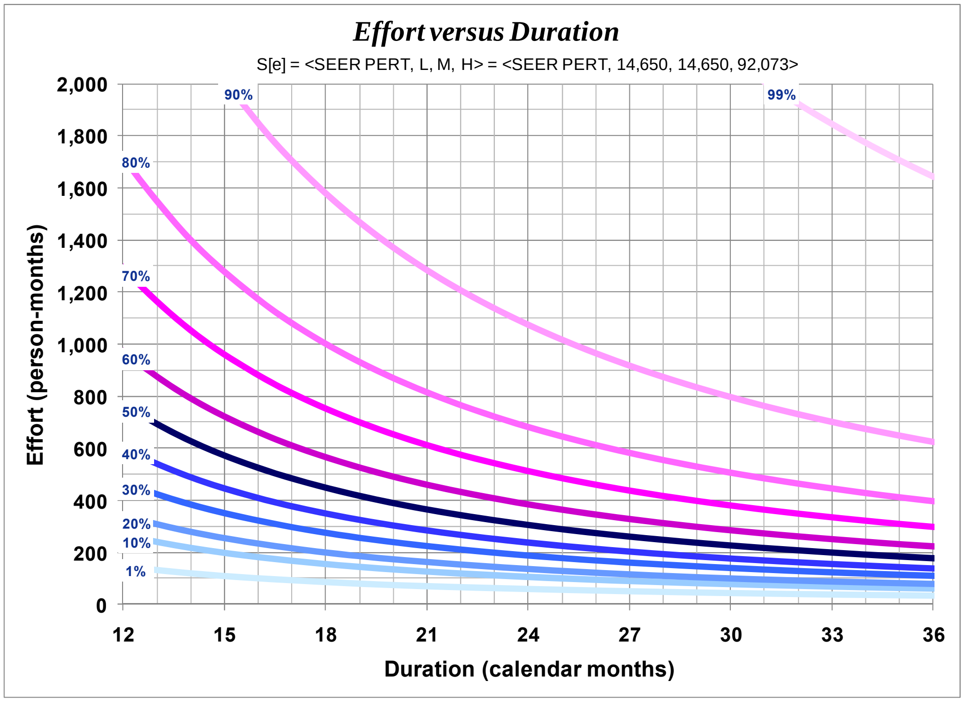
$$\mathbf{\Psi}_{\mathbf{CDF}} = \begin{bmatrix} \psi_1 & p_1 \\ \psi_2 & p_2 \\ \vdots & \vdots \\ \psi_M & p_M \end{bmatrix} \quad \textbf{(47)}$$

Work and Intensity Confidence Level (Attainment Probability) as Fields

Recall that the relationship between our two dependent variables effort and duration is based on the expected size*difficulty product (expected work $\boldsymbol{\Psi}$) which we now treat as a random variable. We begin by refining the definition of our work relationship curve shown in Figure 2 to be a probability field as shown in Figure 3 below.

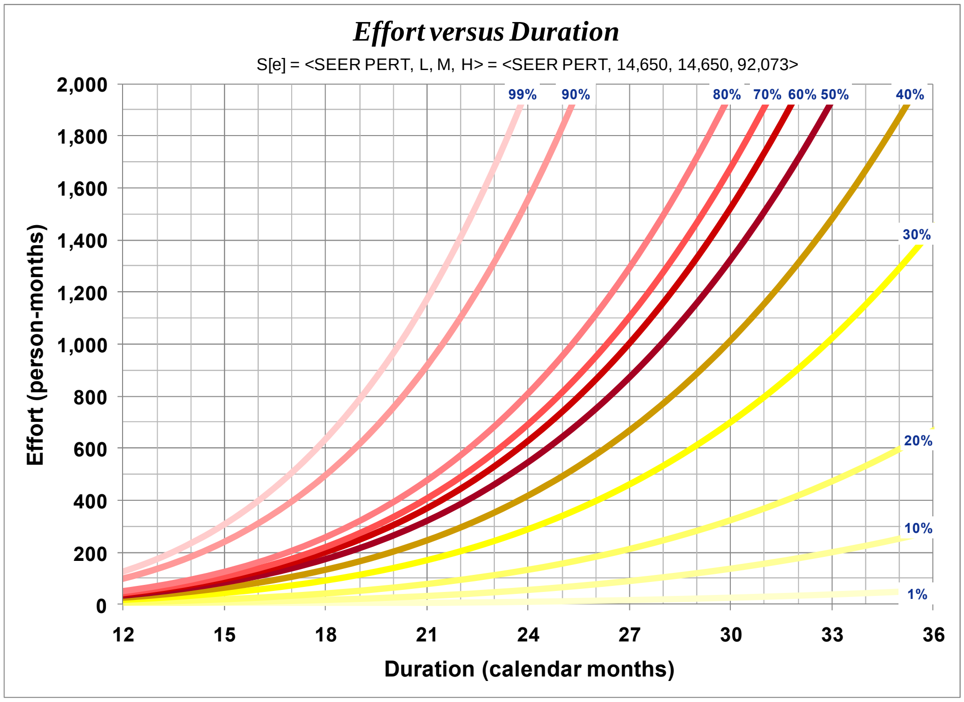
[7] Several algorithms and tools are publically available for doing this including Microsoft Excel's *Percentile* function.

Figure 3: *Work Function as a Probability Field*



We can do the same thing with the intensity relationship as is shown in Figure 4 below.

Figure 4: *Intensity Function as a Probability Field*



We can overlay Figure 3 and Figure 4, as is shown in Figure 5 below, to see how these two functions interact and to get a feel for the estimating scenario's solution space (range of possible outcomes with associated confidence levels (attainment probabilities).

Figure 5: *Work and Intensity Probability Fields*

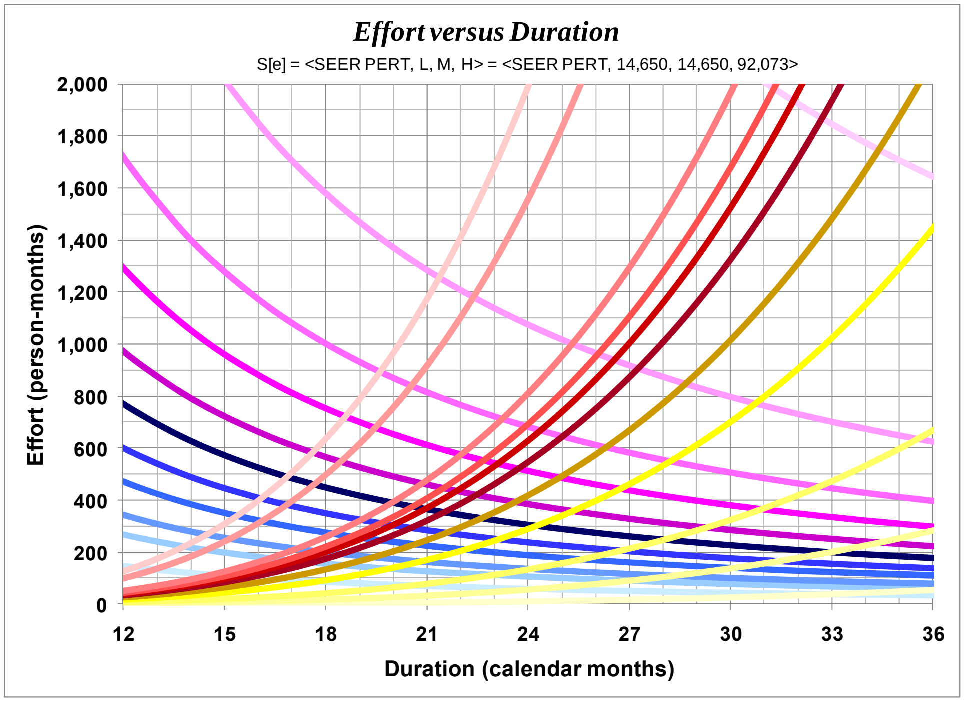


Figure 3, Figure 4, and Figure 5 above show specific members (decade probability field lines) of the represented field. The mathematical expressions for describing any work or intensity field line are, for work

Work

$$\boldsymbol{E}\left(p\,|\,T\right)=\left(\left(\boldsymbol{F}_{\boldsymbol{\Psi}}^{-1}\left(p\right)\right)T^{-\alpha_T}\right)^{\frac{1}{\alpha_E}} \quad \text{and} \quad \boldsymbol{T}\left(p\,|\,E\right)=\left(\left(\boldsymbol{F}_{\boldsymbol{\Psi}}^{-1}\left(p\right)\right)E^{-\alpha_E}\right)^{\frac{1}{\alpha_T}} \qquad \textbf{(48)}$$

Intensity

$$\boldsymbol{E}\left(p\,|\,T\right)=\left(\boldsymbol{F}_{\boldsymbol{I}}^{-1}\left(p\right)\right)T^{\gamma} \quad \text{and} \quad \boldsymbol{T}\left(p\,|\,E\right)=\left(\frac{E}{\left(\boldsymbol{F}_{\boldsymbol{I}}^{-1}\left(p\right)\right)}\right)^{\frac{1}{\gamma}} \qquad \textbf{(49)}$$

where

$\boldsymbol{\Psi} \equiv$ work random variable

$\boldsymbol{I} \equiv$ intensity random variable

$\boldsymbol{F}_{\boldsymbol{\Psi}}^{-1}\left(p\right) \equiv$ inverse work CDF (quantile function) of random variable $\boldsymbol{\Psi}$ at probability p

$\boldsymbol{F}_{\boldsymbol{I}}^{-1}\left(p\right) \equiv$ inverse intensity CDF (quantile function) of random variable $\boldsymbol{I}$ at probability p

Joint Confidence Level (JCL)

Joint Confidence Level (JCL), also known as joint probability, is simply the probability or likelihood that two or more events will occur simultaneously. Suppose we have two events, actual cost being less than or equal to predicted cost and actual schedule being less than or equal to predicted schedule. Since it is desirable that both these events turn out to be true, we might like to know, in addition to the individual probabilities of occurrence, the probability that both will occur. Expressed mathematically

$$\text{JCL (Joint Probability)} \equiv \boldsymbol{P}(A, B) \text{ or } \boldsymbol{P}(A \wedge B) \tag{50}$$

where A and B represent the occurrence of the two events.

We can represent actual cost and schedule as random variables $\boldsymbol{E}$ and $\boldsymbol{T}$ respectively since their outcomes are uncertain; i.e., there is some range of possible outcomes. We treat predicted (estimated) cost and schedule as given specific values $\hat{E}$ and $\hat{T}$ respectively. We rewrite the JCL Equation (50) with these variables as

$$P\left(\left(\boldsymbol{E} \le \hat{E}\right) \wedge \left(\boldsymbol{T} \le \hat{T}\right)\right) \tag{51}$$

where

$\wedge \equiv$ Boolean (logical) AND operator

Note that the two events in Equation (51) are each represented as a Boolean expression, the expressions being separated by a Boolean operator. This results in an overall expression that can evaluate to one of only two possible outcomes for a given pair of $\boldsymbol{E}$ and $\boldsymbol{T}$ draws, TRUE or FALSE. Therefore, the result of $\left(\boldsymbol{E} \le \hat{E}\right) \wedge \left(\boldsymbol{T} \le \hat{T}\right)$ is a random variable we will call $\boldsymbol{J}$ that can be modeled as a discrete distribution of TRUE and FALSE (0 and 1) values. We can define a CDF $\boldsymbol{F_J}$ on the random variable $\boldsymbol{J}$ such that

$$\boldsymbol{F_J}(x \mid \boldsymbol{J}) \equiv \begin{cases} 1 - \dfrac{\sum \boldsymbol{J}}{M} & x = 0 \text{ (FALSE)} \\ \dfrac{\sum \boldsymbol{J}}{M} & x = 1 \text{ (TRUE)} \end{cases} \tag{52}$$
$$x \in \{0, 1\}$$

Conditional Confidence Level (CCL)

Conditional Confidence Level (CCL), also known as conditional probability, is simply the probability or likelihood of some event given the occurrence of some other event. Suppose we have two events, actual cost being less than or equal to predicted cost and an assumption that actual schedule will equal the predicted schedule. Conversely we could have two events, actual schedule being less than or equal to predicted schedule and an assumption that actual cost will equal the predicted cost. Expressed mathematically

$$\text{CCL (Conditional Probability)} \equiv \boldsymbol{P}(A \mid B = 1 \text{ (TRUE)}) \tag{53}$$

or

$$\text{CCL (Conditional Probability)} \equiv \boldsymbol{P}\left(B \,|\, A = 1 \text{ (TRUE)}\right) \tag{54}$$

where A and B represent the occurrence of the two events.

As with JCL, we can represent actual cost and schedule as random variables $\boldsymbol{E}$ and $\boldsymbol{T}$ respectively since their outcomes are uncertain; i.e., there is some range of possible outcomes. We treat predicted (estimated) cost and schedule as given specific values $\hat{E}$ and $\hat{T}$ respectively. We re-write the CCL Equations (53) and (54) with these variables as

$$\boldsymbol{P}\left(\boldsymbol{E} \le \hat{E} \,|\, \hat{T}\right) \tag{55}$$

or

$$\boldsymbol{P}\left(\boldsymbol{T} \le \hat{T} \,|\, \hat{E}\right) \tag{56}$$

A common requirement of cost analysts is the ability to perform what-if or excursion analysis from some baseline estimating scenario. Performing these kinds of analyses generally implies the use of some form of conditional probability; i.e., solving for one or more variables conditional on some other variable(s) being assumed to take on some specific value(s). Our defined system of correlated estimating relationships offers the opportunity to perform these kinds of analyses by taking advantage of the fact that the CER and SER are correlated by intensity. If we treat intensity as having a specific value rather than as a random variable, we can use intensity as a gradient function across the work probability field. Practical application of this method is shown in the examples found later in the paper. Equations (57) and (58) below provide the means for calculating an intensity value that corresponds to a particular effort or duration value at a particular confidence level.

Finding the Intensity that Satisfies a Cost (Effort) Constraint

$$\left[I_{effort\_constraint@p\%} = \left(\frac{E_{effort\_constraint@p\%}^{\gamma\alpha_E + \alpha_T}}{\boldsymbol{F}_{\boldsymbol{\Psi}}^{-1}\left(p\right)^{\gamma}} \right)^{1/\alpha_T} \right]_{\text{<dataset name>}} \tag{57}$$

where

$\boldsymbol{F}_{\boldsymbol{\Psi}}^{-1}\left(p\right) \equiv$ inverse CDF of random variable $\boldsymbol{\Psi}$ at probability p

Finding the Intensity that Satisfies a Schedule (Duration) Constraint

$$\left[I_{duration\_constraint@p\%} = \left(\frac{\boldsymbol{F}_{\boldsymbol{\Psi}}^{-1}\left(p\right)}{T_{duration\_constraint@p\%}^{\gamma\alpha_E + \alpha_T}} \right)^{1/\alpha_E} \right]_{\text{<dataset name>}} \tag{58}$$

5. Examples

As a way to demonstrate some of the analyses that can be performed with the above-described model we attempt to answer the three questions posed in Dr. Paul Garvey's (2000) challenge. We do this with a series of graphs followed by the mathematics that support the graphs.

Response to Garvey Question 1

"What is the chance the system can be delivered within cost and schedule?"

We assume this question is seeking the JCL (joint cost and schedule probability) of some given single-point cost and schedule position. Figure 6, Figure 7, and Figure 8 show 2001 Monte Carlo draws from the earlier-regressed CDER elaborated with a SEER PERT distributed size of $\langle Least, Likely, Most \rangle = \langle 14650, 14650, 92073 \rangle$. For this example we assume the given single-point estimate of cost (effort) and of schedule (duration) in the stated question to be the values at the intersection of the mean work field curve located by $\bar{\Psi}$ and the mean intensity field curve located by $\bar{I}$. This assumption is arbitrary, we could have assumed any effort-duration pair in the solution space to represent the budget, goal, constraint, etc. since this position is not stated in the question. What matters here is the methodology for answering this kind of question.

Graphical Solution

Figure 6: *Scatter Diagram of Monte Carlo Draws*

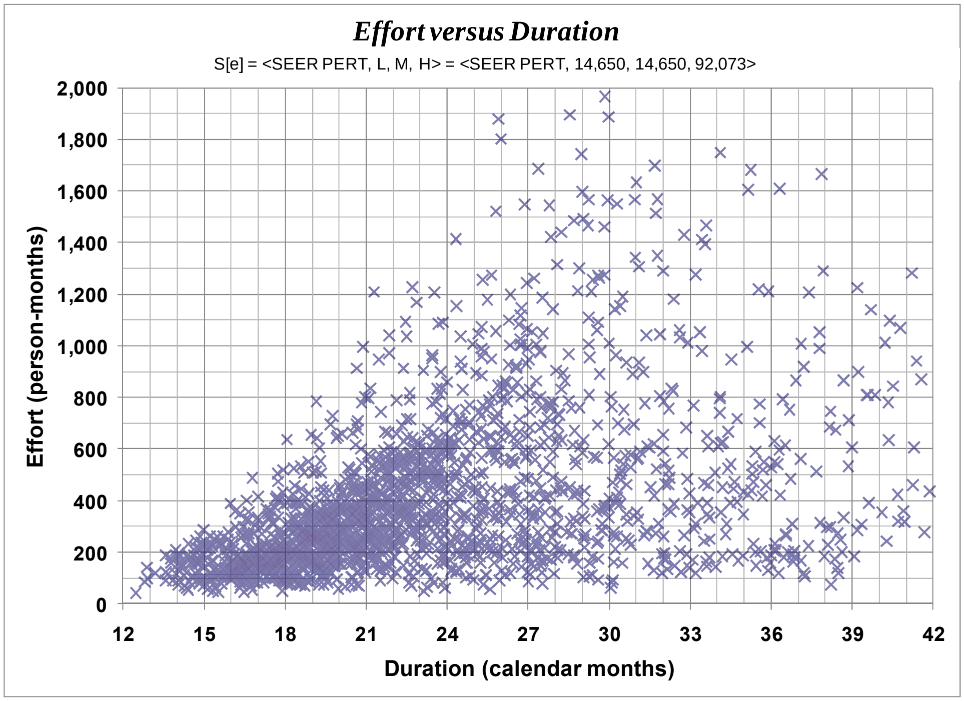


Figure 7: *Scatter Diagram of Monte Carlo Draws with Point Estimate Position*

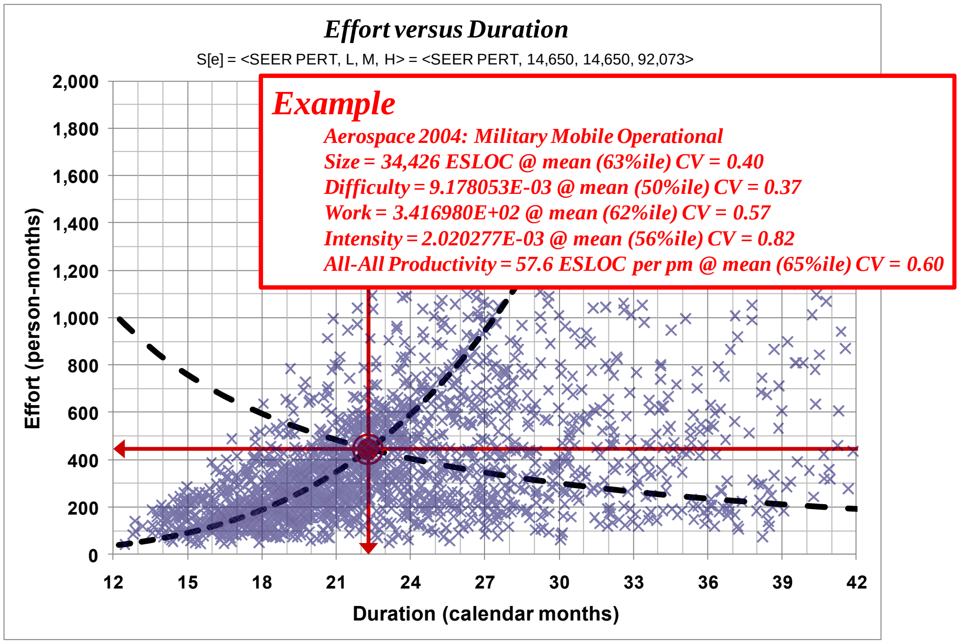
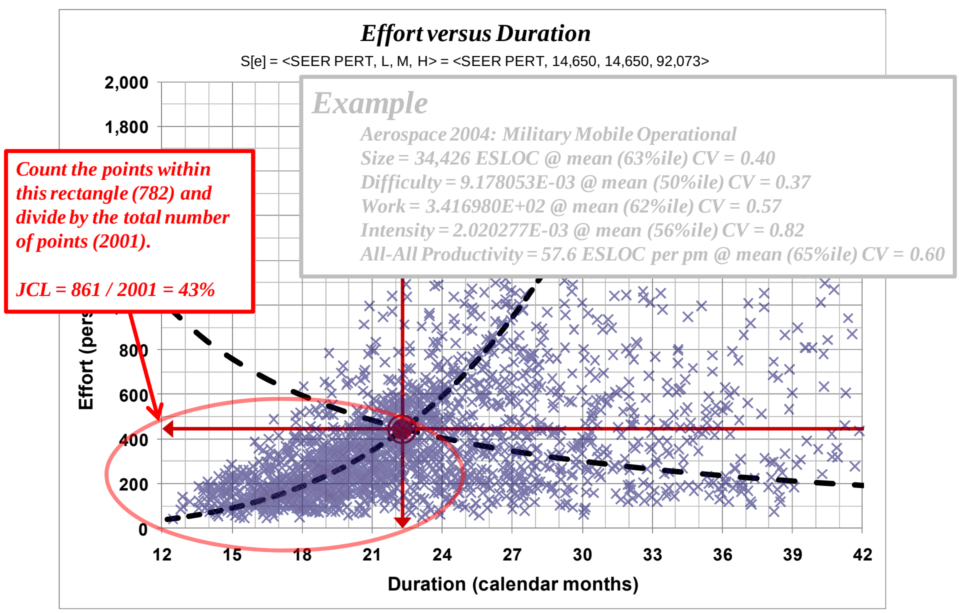


Figure 8: *Illustration of Area where Monte Carlo Draws Satisfy Both Constraints*



Since it is not practical to actually count points on a scatter diagram, the mathematics shown below applied to vectors of draws from the relevant random variables will yield the solution.

$$\boldsymbol{J} \equiv \left(\boldsymbol{E} \le \hat{E}\right) \wedge \left(\boldsymbol{T} \le \hat{T}\right)$$

$$\boldsymbol{F}_{\boldsymbol{J}}\left(x \mid \boldsymbol{J}\right) \equiv \begin{cases} 1 - \dfrac{\sum \boldsymbol{J}}{2001} & x = 0 \ \left(\text{FALSE}\right) \\ \dfrac{\sum \boldsymbol{J}}{2001} & x = 1 \ \left(\text{TRUE}\right) \end{cases} \tag{59}$$

$$\boldsymbol{F}_{\boldsymbol{J}}\left(\text{TRUE}\right) = 861/2001 = 43\%$$

$$\therefore \text{Joint Confidence Level} = 43\%$$

Response to Garvey Question 2

"How likely might the point estimate cost be exceeded for a given schedule?"

This question suggests a conditional probability excursion. As in the previous example we assume the point estimate cost (effort) and schedule (duration) in the stated question to be the values at the intersection of the mean work field curve located by $\bar{\Psi}$ and the mean intensity field curve located by $\bar{I}$. Again, this assumption is arbitrary, we could have assumed any effort-duration pair in the solution space.

Graphical Solution

We start by locating the point estimate position

Figure 9: *Locate Point Estimate Position*

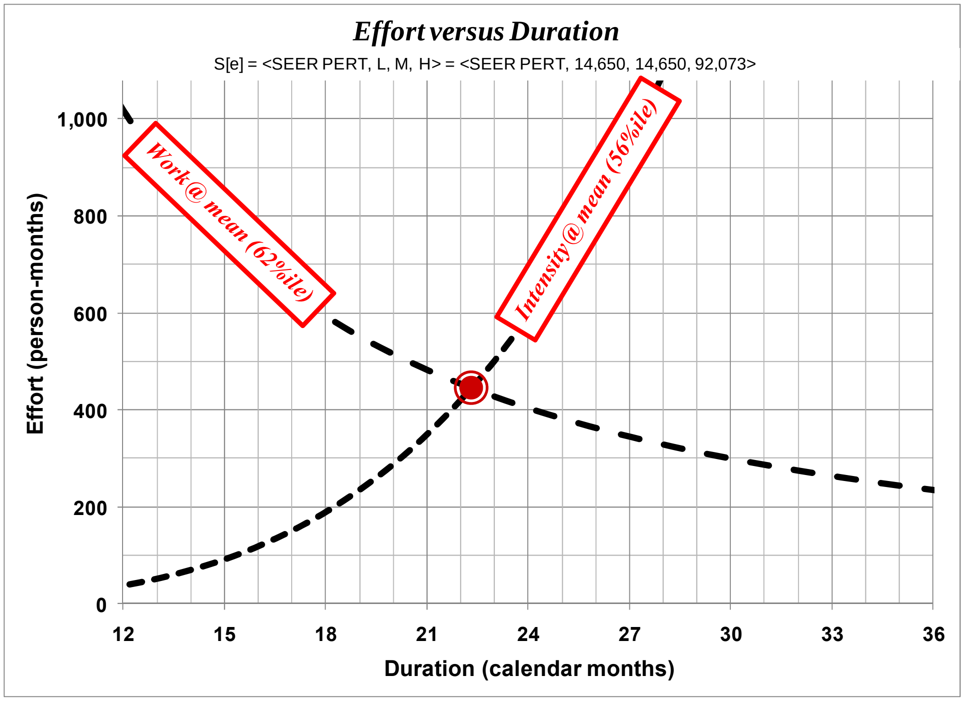


Figure 10: *Point Estimate Values*

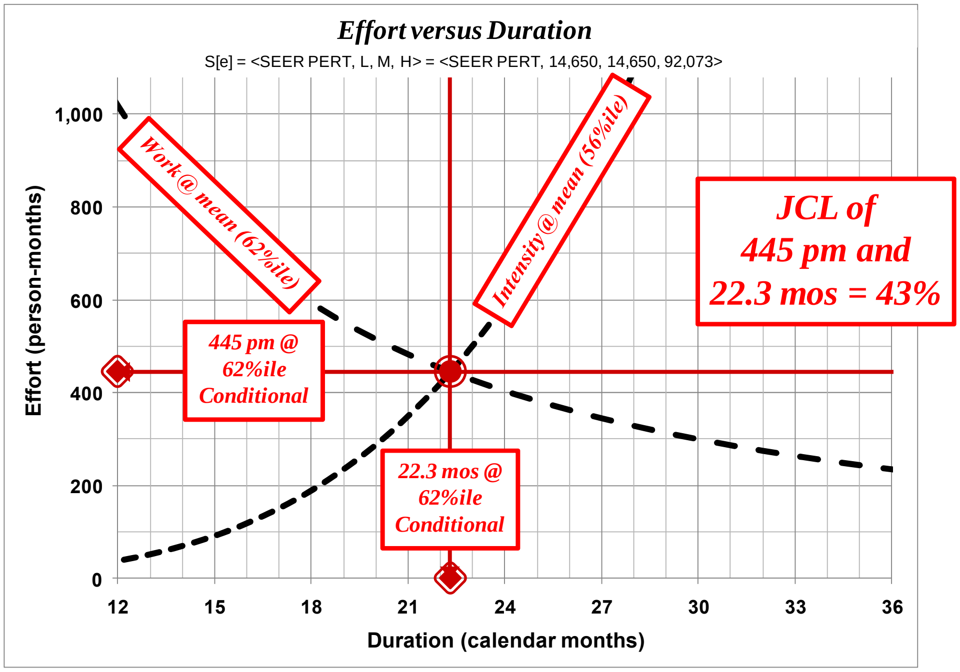


Figure 11: *Establish Updated Schedule Constraint*

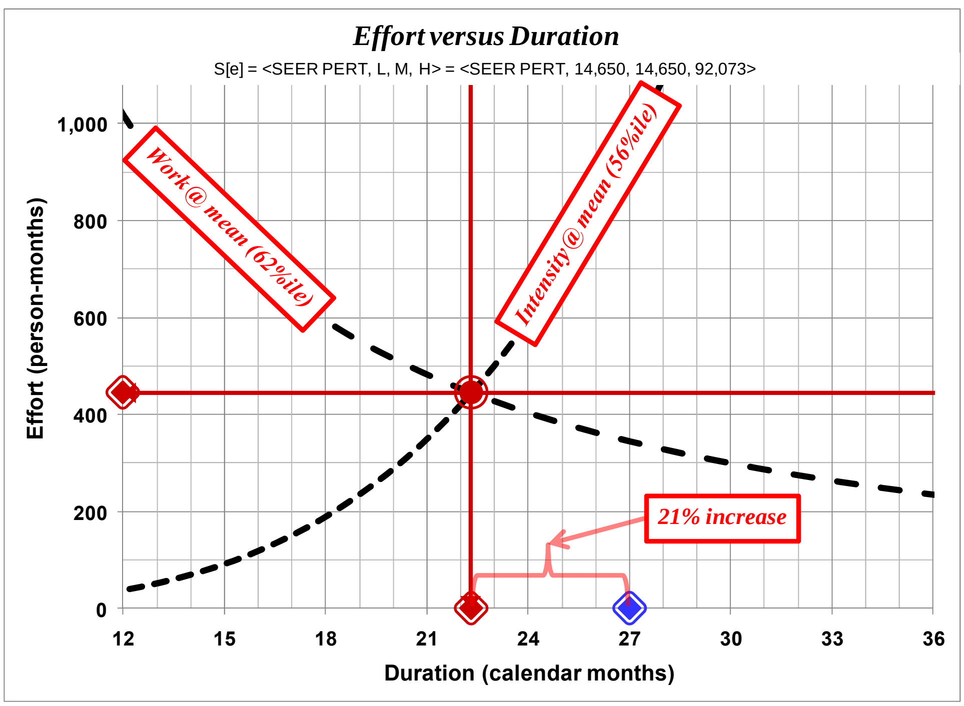


Figure 12: *Locate Position for Intensity Update*

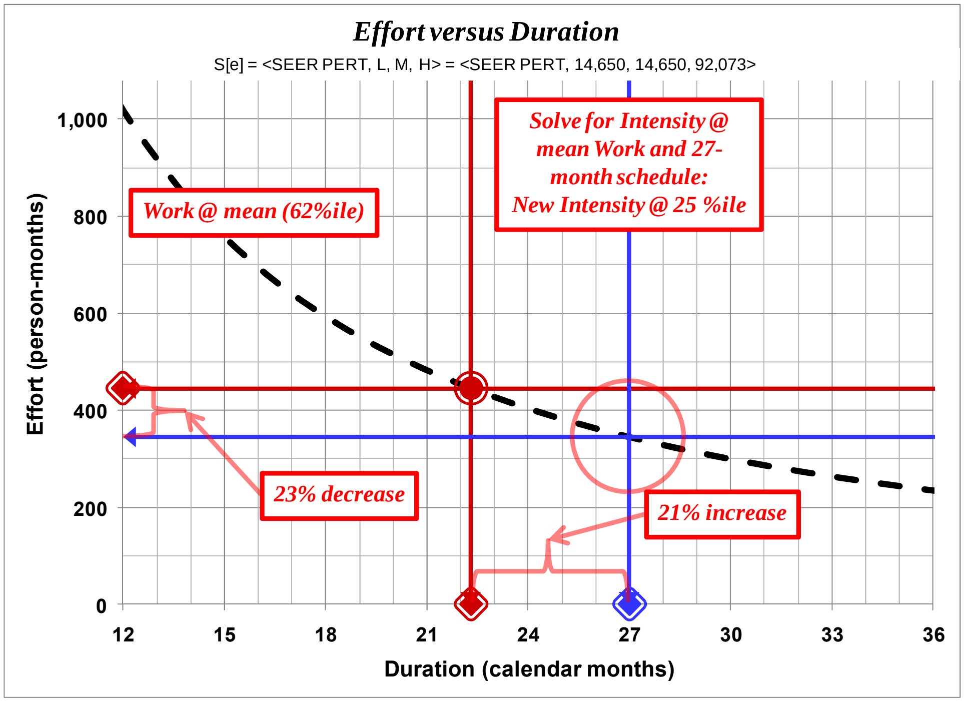


Figure 13: *Update Intensity Location*

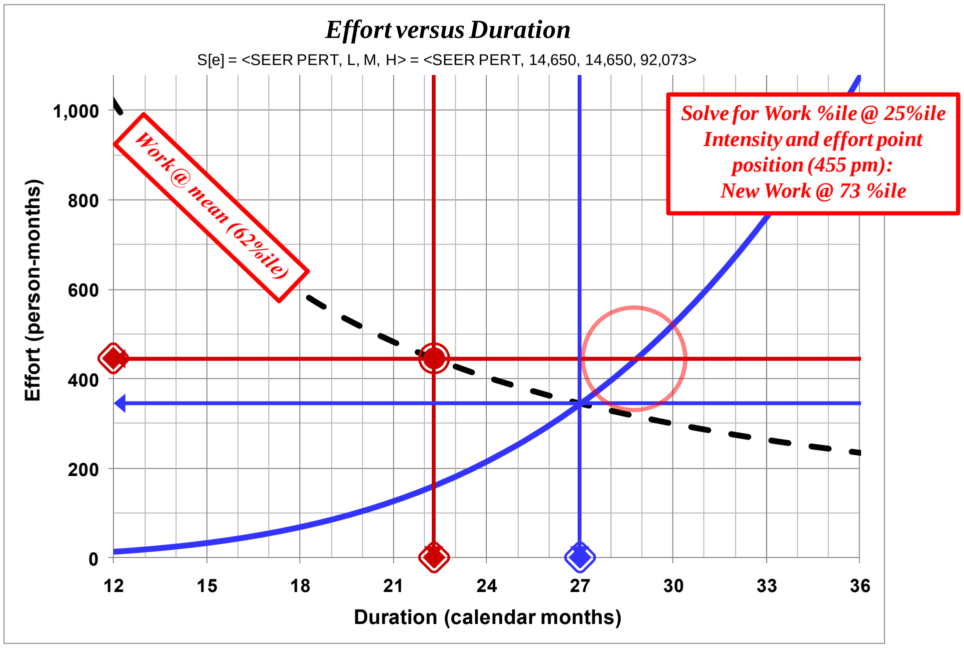
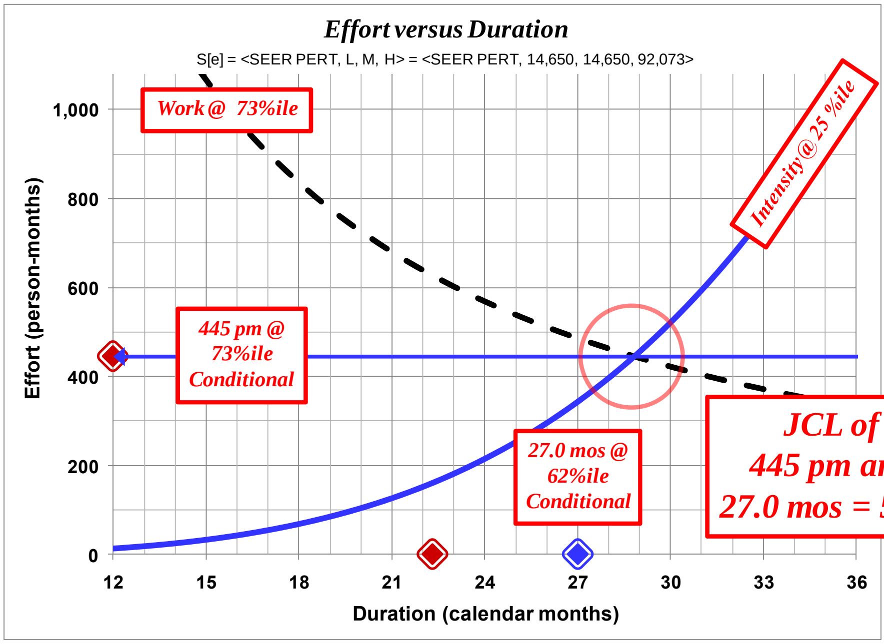


Figure 14: *Solution*



Notice how the intensity curve is being used as a gradient function across the work field to find the appropriate work field curve that maintains an intensity value consistent with a 27 month schedule with a confidence level of 62% while satisfying the effort value. The confidence level of the discovered work field curve (73%) is the new confidence level of the 445 person-month effort value.

Mathematical Solution

Assume the point estimate position to be the effort-duration pair that is defined by $\bar{\Psi}$ and $\bar{I}$. Locate the new intensity I_1 gradient and then use it to populate the new effort random variable $\boldsymbol{E}_{I_1}$.

$$T_{point} \equiv 22.3 \quad p_{T_{point}} \equiv 52\% \quad T_1 \equiv 27.0 \quad p_{T_1} \equiv 52\% \quad E_{point} \equiv 445 \quad p_{E_{Point}} \equiv 52\%$$

$$\left[I_1 = \left(\frac{\boldsymbol{F}_{\boldsymbol{\Psi}}^{-1}\left(p_{T_1}\right)}{T_1^{\gamma\alpha_E+\alpha_T}} \right)^{1/\alpha_E} = \left(\frac{\boldsymbol{F}_{\boldsymbol{\Psi}}^{-1}(52\%)}{27.0^{3.96(0.57+0.76)}} \right)^{1/0.57} = 7.321961\text{E}-04 \right]_{<*>} \qquad \textbf{(60)}$$

$$\left[\boldsymbol{E}_{I_1} = \left(I_1^{\alpha_T} \boldsymbol{\Psi}^{\gamma} \right)^{1/(\gamma\alpha_E+\alpha_T)} = \left(I_1^{0.76} \boldsymbol{\Psi}^{3.96} \right)^{1/(3.96(0.57+0.76))} \right]_{<*>}$$

$$\therefore \boldsymbol{P}\left(\boldsymbol{E}_{I_1} > E_{point}\right) = 1 - \boldsymbol{F}_{\boldsymbol{E}_{I_1}}\left(E_{point}\right) = 1 - \boldsymbol{F}_{\boldsymbol{E}_{I_1}}(445) = 1 - 73\% = 27\%$$

where

$F_X(x) \equiv$ CDF of random variable X at value x

$F_X^{-1}(p) \equiv$ inverse CDF of random variable X at probability p

<*> indicates Aerospace 2004: Military Ground Operational

Response to Garvey Question 3

"How are cost reserve recommendations affected by schedule risk?"

Graphical Solution

Figure 15: *Locate Appropriate Work Curve*

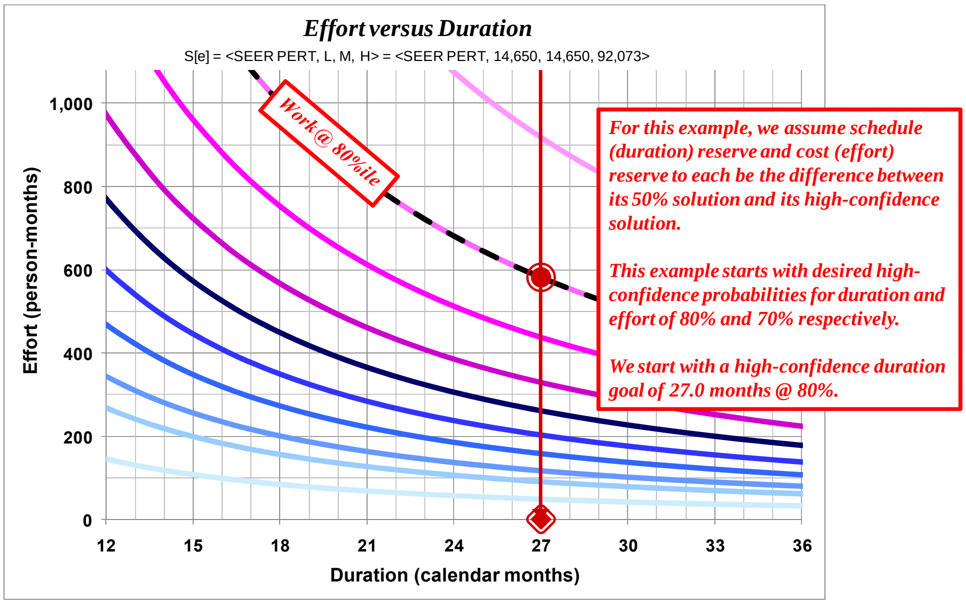
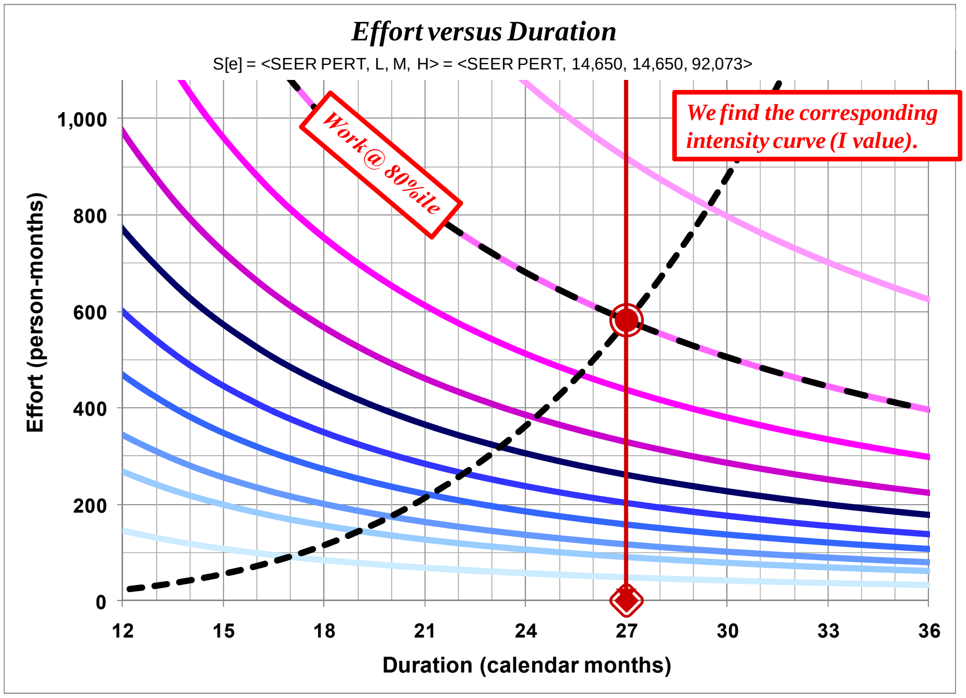
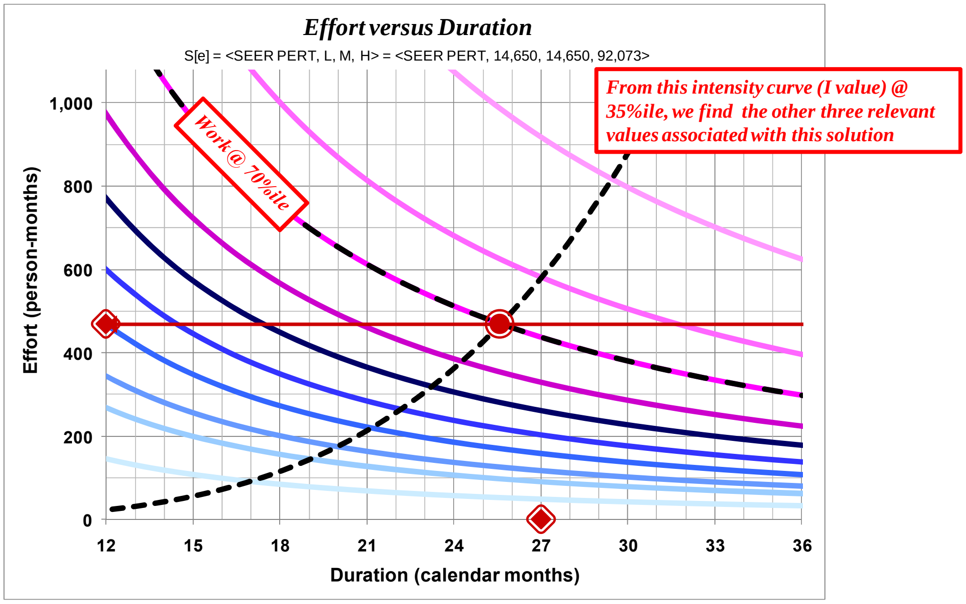


Figure 16: *Locate Corresponding Intensity Curve*



Notice once again how the intensity curve is being used as a gradient function across the work field, in this case to find the intensity curve that corresponds to a duration value of 27 months at a confidence level of 80%.

Figure 17: *Locate and Project Constrained Effort Solution*



The discovered intensity curve is used here to *orthographically project* effort with a 70% confidence level, effort with a 50% confidence level, and duration with a 50% confidence level.

Figure 18: *Locate and Project Constrained 50% Confidence Solutions*

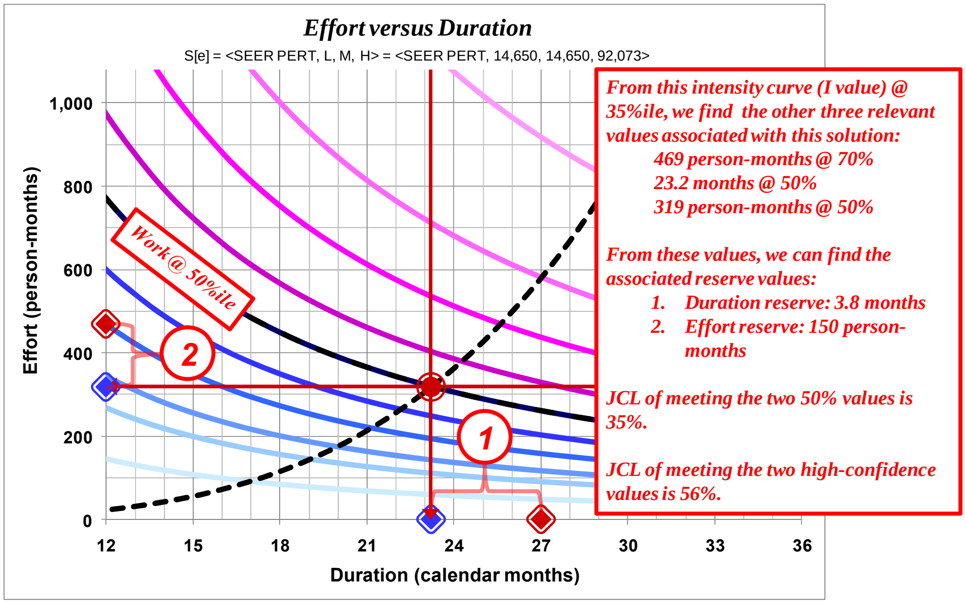
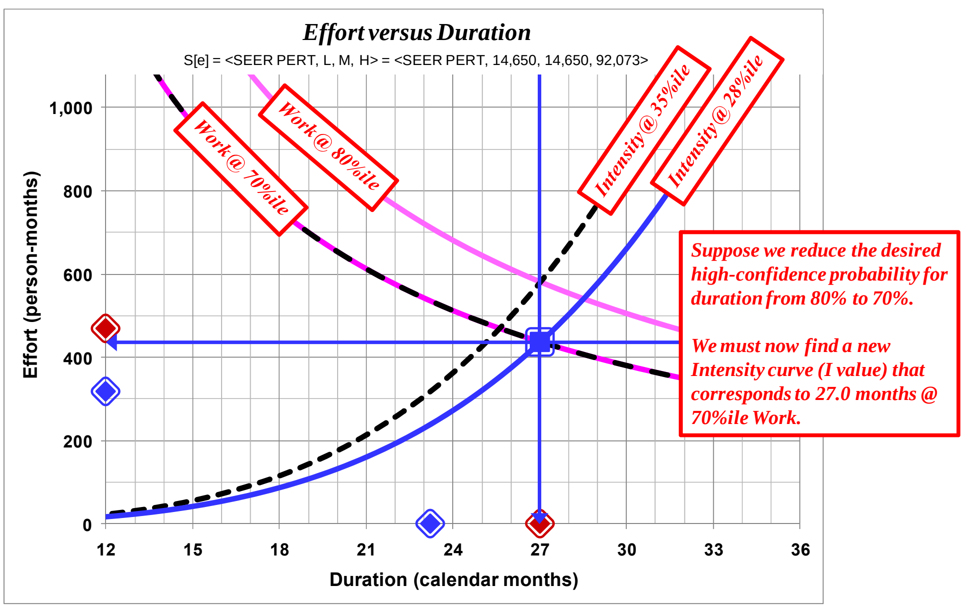
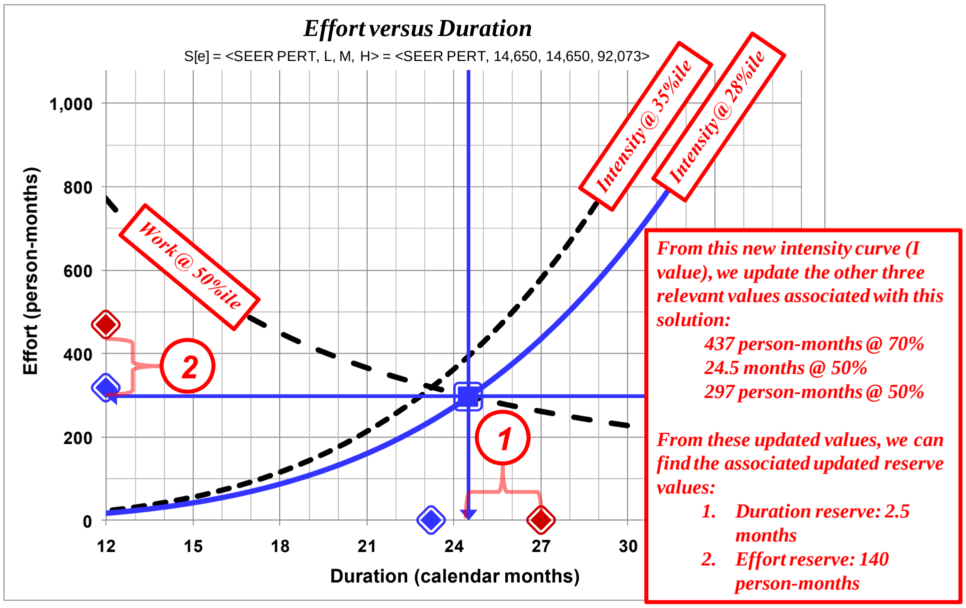


Figure 19: *Update Intensity Curve Location*



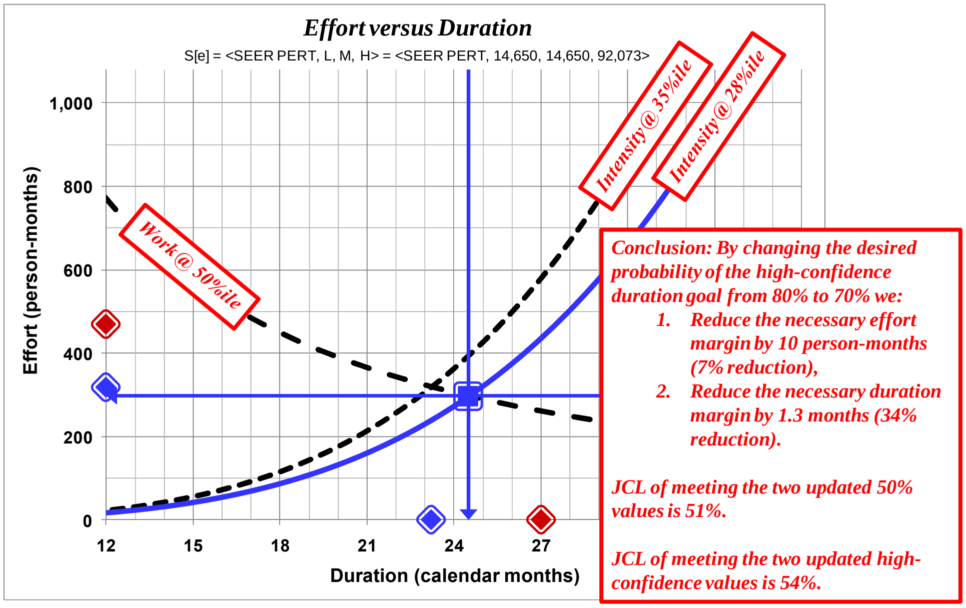
Note that a new intensity curve has been discovered that corresponds with the 27 month schedule constraint at an updated confidence level of 70% (down from the original 80%).

Figure 20: *Project Updated Effort and Duration Positions*



Once again the discovered intensity curve is used as a gradient function across the work field to locate and orthographically project updated effort and duration values at the originally-given confidence levels.

Figure 21: *Solution*



Mathematical Solution

$$T_{1\_High\_Confidence} = T_{2\_High\_Confidence} \equiv 27.0$$
$$p_{T_{1\_High\_Confidence}} \equiv 80\% \quad p_{T_{2\_High\_Confidence}} \equiv 70\% \tag{61}$$
$$p_{E_{1\_High\_Confidence}} = p_{E_{2\_High\_Confidence}} \equiv 70\%$$

$$\left[I_1 \equiv \left(\frac{\boldsymbol{F}_{\boldsymbol{\Psi}}^{-1}\left(p_{T_{1\_High\_Confidence}}\right)}{T_{1\_High\_Confidence}^{\ \gamma\alpha_E+\alpha_t}} \right)^{1/\alpha_E} = \left(\frac{\boldsymbol{F}_{\boldsymbol{\Psi}}^{-1}(80\%)}{27.0^{3.96(0.57+0.76)}} \right)^{1/0.57} = 1.235254\text{E}-03 \right]_{<*>} \tag{62}$$

where

$\boldsymbol{F}_{\boldsymbol{X}}(x) \equiv$ CDF of random variable $\boldsymbol{X}$ at value x

$\boldsymbol{F}_{\boldsymbol{X}}^{-1}(p) \equiv$ inverse CDF of random variable $\boldsymbol{X}$ at probability p

<*> indicates Aerospace 2004: Military Ground Operational

$$\left[\boldsymbol{E}_{I_1} \equiv \left(I_1^{\alpha_T} \boldsymbol{\Psi}^{\gamma} \right)^{1/(\gamma\alpha_E+\alpha_T)} = \left((1.235254\text{E}-03)^{0.76} \boldsymbol{\Psi}^{3.96} \right)^{1/(3.96(0.57+0.76))} \right]_{<*>}$$
$$E_{1\_High\_Confidence} \equiv \boldsymbol{F}_{\boldsymbol{E}_{I_1}}^{-1}\left(p_{E_{1\_High\_Confidence}}\right) = \boldsymbol{F}_{\boldsymbol{E}_{I_1}}^{-1}(70\%) = 469 \tag{63}$$
$$E_{1\_50\%} \equiv \boldsymbol{F}_{\boldsymbol{E}_{I_1}}^{-1}(50\%) = 319$$

$$\left[\boldsymbol{T}_{I_1} = \left(\frac{1}{I_1^{\alpha_E}} \boldsymbol{\Psi} \right)^{1/(\gamma\alpha_E+\alpha_T)} = \left(\frac{1}{(1.235254\text{E}-03)^{0.57}} \boldsymbol{\Psi} \right)^{1/(3.96(0.57+0.76))} \right]_{<*>} \tag{64}$$

$$T_{1\_50\%} \equiv \boldsymbol{F}_{\boldsymbol{T}_{I_1}}^{-1}\left(p_{T_{1\_High\_Confidence}}\right) = \boldsymbol{F}_{\boldsymbol{T}_{I_1}}^{-1}(50\%) = 23.2$$
$$\Delta_{T_1} = T_{1\_High\_Confidence} - T_{1\_50\%} = 27.0 - 23.2 = 3.8 \tag{65}$$
$$\Delta_{E_1} = E_{1\_High\_Confidence} - E_{1\_50\%} = 469 - 319 = 150$$

$$\left[I_2 = \left(\frac{\boldsymbol{F}_{\boldsymbol{\Psi}}^{-1}\left(p_{t_{2\_High\_Confidence}}\right)}{T_{2\_High\_Confidence}^{\ \gamma\alpha_E+\alpha_T}} \right)^{1/\alpha_E} = \left(\frac{\boldsymbol{F}_{\boldsymbol{\Psi}}^{-1}(70\%)}{36.0^{3.96(0.57+0.76)}} \right)^{1/0.57} = 9.287725\text{E}-04 \right]_{<*>} \tag{66}$$

$$\left[\boldsymbol{E}_{I_2} \equiv \left(I_2^{\alpha_T} \boldsymbol{\Psi}^{\gamma} \right)^{1/(\gamma\alpha_E+\alpha_T)} = \left((9.287725\text{E}-04)^{0.76} \boldsymbol{\Psi}^{3.96} \right)^{1/(3.96(0.57+0.76))} \right]_{<*>}$$
$$E_{2\_High\_Confidence} \equiv \boldsymbol{F}_{\boldsymbol{E}_{I_2}}^{-1}\left(p_{E_{2\_High\_Confidence}}\right) = \boldsymbol{F}_{\boldsymbol{E}_{I_2}}^{-1}(70\%) = 437 \tag{67}$$
$$E_{2\_50\%} \equiv \boldsymbol{F}_{\boldsymbol{E}_{I_2}}^{-1}(50\%) = 297$$

$$\left[\boldsymbol{T}_{\boldsymbol{M_2}} = \left(\frac{1}{I_2^{\alpha_E}} \boldsymbol{\Psi} \right)^{1/(\gamma\alpha_E + \alpha_T)} = \left(\frac{1}{(9.287725\text{E}-04)^{0.57}} \boldsymbol{\Psi} \right)^{1/(3.96(0.57+0.76))} \right]_{<*>} \quad \textbf{(68)}$$

$$T_{2\_50\%} \equiv \boldsymbol{F}_{\boldsymbol{T_{M_2}}}^{-1}\left(p_{T_{2\_High\_Confidence}} \right) = \boldsymbol{F}_{\boldsymbol{T_{M_2}}}^{-1}(50\%) = 24.5$$

$$\begin{aligned} \Delta_{T_2} &= T_{2\_High\_Confidence} - T_{2\_50\%} = 27.0 - 24.5 = 2.5 \\ \Delta_{E_2} &= E_{2\_High\_Confidence} - E_{2\_50\%} = 437 - 297 = 140 \\ \Delta_T &= 2.5 - 3.8 = -1.3 \quad \Delta_{T\%} = \frac{2.5 - 3.8}{3.8} = -34\% \\ \Delta_E &= 140 - 150 = -10 \quad \Delta_{E\%} = \frac{140 - 150}{150} = -7\% \end{aligned} \quad \textbf{(69)}$$

REFERENCES

Aerospace Corporation. 2004. Software Cost and Productivity Model. *Aerospace Report No. ATR-2004(8311)-1.* El Segundo, CA, USA : s.n., 2004.

Bernstein, Peter L. 1996. *Against the Gods: The Remarkable Story of Risk.* New York City : John Wiley & Sons, Inc., 1996. ISBN 0-471-29563-9 (paper).

Druker, Eric and Anderson, Timothy P. 2009. Joint Cost and Schedule Risk Analysis Guidelines. Canoga Park, CA, USA : UNPUBLISHED DRAFT, September 16, 2009.

Garvey, Paul R. 2000. *Probability Methods for Cost Uncertainty Analysis: A Systems Engineering Perspective.* Boca Raton, London, New York NY : Chapman-Hall/CRC Press, 2000.

National Aeronautics and Space Administration (NASA). 2004. *2004 NASA Cost Estimating Handbook (CEH).* Washington : NASA, 2004. Online Handbook: http://ceh.nasa.gov.

Ross, Michael A. 2011a. An Improved Method for Predicting Software Code Growth: Tecolote DSLOC Estimate Growth Model. *Proceedings, Joint ISPA / SCEA 2011 Conference and Training Workshop.* Albuquerque, NM, USA : The International Society of Parametric Analysts and The Society of Cost Estimating and Analysis, June 2011a.

—. 2007a. Illustrating Probability in Software Cost and Schedule Estimating: Know the Odds Before Placing Your Bet. *Proceedings, AIAA SPACE 2007 Conference & Exhibition.* Long Beach, CA, USA : American Institute of Aeronautics and Astronautics, September 2007a. AIAA 2007-6022.

—. 2008. Next Generation Software Estimating Framework: 25 Years and Thousands of Projects Later. [ed.] Stephen A. Book and Edward White III. *Journal of Cost Analysis and Parametrics.* s.l. : Society of Cost Estimating and Analysis - International Society of Parametric Analysts, Fall 2008. Vol. 1, 2, pp. 7-30. ISSN 1941-658X.

Sambur, Marvin R. and Teets, Peter B. 2004. Revitalizing the Software Aspects of Systems Engineering. *Memorandum for See Distribution.* Washington, District of Columbia, USA : Under Secretary of the Air Force, September 20, 2004. 04A-003.

BIOGRAPHY

Michael A. Ross *has over 35 years of experience in software engineering as a developer, manager, process expert, consultant, instructor, and award-winning international speaker. Mr. Ross is currently a Technical Expert for Tecolote Research, Inc. Mr. Ross's previous experience includes three years as President and CEO of r2Estimating, LLC (makers of the r2Estimator software estimation tool), three years as Chief Engineer of Galorath Inc. (makers of the SEER suite of estimation tools), seven years with Quantitative Software Management, Inc. (makers of the SLIM suite of software estimating tools) where he was Vice President of Education Services, and 17 years with Honeywell Air Transport Systems (formerly Sperry Flight Systems) and 2 years with Tracor Aerospace where he developed and/or managed the development of various military and commercial avionics systems. Mr. Ross did his undergraduate work at the United States Air Force Academy and Arizona State University, receiving a Bachelor of Science in Computer Engineering.*