A Probabilistic Method for Predicting Software Code Growth: 2018 Update

ERIC M. SOMMER
USAF Space Command SMC/FMC, El Segundo, California

BOPHA SENG
Tecolote Research, Inc., El Segundo, California

DAVID L. LAPORTE
Tecolote Research, Inc., El Segundo, California

MICHAEL A. ROSS
r2Estimating, LLC, Lakeside, Montana

ABSTRACT

Software estimating is challenging. SMC's approach has evolved over time to tackle this challenge. Originally based on Mike Ross's 2011 DSLOC Estimate Growth Model, we've updated our model to include more recent SRDR data and an improved methodology (Orthogonal Distance Regression). Discussions will focus on non-linear relationships between size and growth, unique growth for new, modified, and unmodified DSLOC, as well as correlation between DSLOC types and future efforts to include space flight software data.

1. INTRODUCTION

The Delivered Source Lines Of Code (DSLOC) Estimate Growth Model version 8 (DEGM8) provides probabilistic growth adjustments to single-point Technical Baseline Estimates (TBEs) of Delivered Source Lines of Code (DSLOC), for New software, Modified software, and Unmodified software, that are sensitive to the estimate maturity of the DSLOC TBEs; i.e., when, in the Software Development Life Cycle (SDLC), the DSLOC TBE is performed. It is a data-driven model and methodology that is based on Software Resources Data Report (SRDR) data collected and archived by the U.S. Department of Defense’s Defense Cost and Resource Center (DCAR). This model represents a significant update and modernization of the Tecolote DSLOC Estimate Growth Model version 7 (DEGM7) (Ross, 2011) in that:

- It is based on recently-updated SRDR data.
- It is based on a better method of regressing the historical data.
- It recognizes non-linear relationships between size and growth.
• It accounts for error (uncertainty) in both the input DSLOC TBEs and the output DSLOC estimates.
• It decomposes the DEGM7 notion of Pre-existing reused software into Modified software and Unmodified software.
• It recognizes correlation between New, Modified, and Unmodified growth.

This new version will be released as DSLOC Estimate Growth Model version 8 (DEGM8).

This paper first summarizes the equations that comprise the model. It then provides a detailed description of the model’s basis and components. Next, it describes an example of how to use the model to calculate growth-adjusted software size estimate distributions for the New, Modified, and Unmodified DSLOC that make up an example Computer Software Configuration Item (CSCI). The paper concludes with the authors’ collective opinion of the value this model represents.

This paper includes:

• A section describing DEGM8SV (a special instance of the DEGM8) that estimates DSLOC associated with unmanned Space Vehicle (SV) flight software,
• An appendix containing Custom Cumulative Distribution Function (CDF) tables that can be copied into tools such as ACEIT or Crystal Ball in order to construct custom CDFs1 that are needed to model the baseline New, Modified, and Unmodified DSLOC error factor parameter distributions,
• An appendix containing the regression results of three different data set filtering alternatives,
• An appendix containing a graphic comparison of DEGM7 and DEGM8 behavior,
• An appendix containing a detailed mathematical description of Orthogonal Distance Regression (ODR), a special case of Total Least Squares Regression, and how it is applied to the SRDR DSLOC data,
• An appendix containing a TRI ACEIT implementation of the paper’s example growth-adjusted DSLOC estimate.
2. MODEL SUMMARY

The DEGM8 equations for applying growth and uncertainty to TBE New, Modified, and Unmodified DSLOC are shown in Figure 1 below.

\[ S_{DGANew} \triangleq S_{DNew} + e^{-(Decay)(Maturity)} \left( \tilde{b}_{GN} \varepsilon_{GN} \left( \frac{S_{DNew}}{K_N} \right)^{a_{GN}} K_N - S_{DNew} \right) \]

\[ S_{DGAMod} \triangleq S_{DMod} + e^{-(Decay)(Maturity)} \left( \tilde{b}_{GM} \varepsilon_{GM} \left( \frac{S_{DMod}}{K_M} \right)^{a_{GM}} K_M - S_{DMod} \right) \]

\[ S_{DGAUmod} \triangleq S_{DUmod} + e^{-(Decay)(Maturity)} \left( \tilde{b}_{GU} \varepsilon_{GU} \left( \frac{S_{DUmod}}{K_U} \right)^{a_{GU}} K_U - S_{DUmod} \right) \]

**Figure 1** DEGM8 equations yield the sum of the appropriate TBE DSLOC value and its calculated DSLOC growth amount. The calculated DSLOC Growth amount is the product of the baseline DSLOC growth amount (zero maturity) and the calculated estimate maturity adjustment factor.

where

- \( S_{DGANew} \) = Output – growth-adjusted New DSLOC distribution of outcomes with associated attainment probability
- \( S_{DGAMod} \) = Output – growth-adjusted Modified DSLOC estimate distribution of outcomes with associated attainment probability
- \( S_{DGAUmod} \) = Output – growth-adjusted Unmodified DSLOC estimate distribution of outcomes with associated attainment probability
- \( \triangleq \) = Estimator equality symbol; the right expression estimates the left expression
- \( S_{DNew} \) = Input – Technical Baseline Estimate (TBE) of New DSLOC
- \( S_{DMod} \) = Input – Technical Baseline Estimate (TBE) of Modified DSLOC

The following section of this paper describes the basis of these equations.

3. COMPONENTS OF THE MODEL

The DEGM8 estimates, as shown in Figure 1 above, *growth-adjusted total amounts for New, Modified, and Unmodified DSLOC* that are each calculated as the sum of the particular *TBE DSLOC amount* and its associated *maturity-adjusted DSLOC growth amount*. 
\[ S_{DGANew} \triangleq S_{DNew} + f(Maturity)S_{DGAmountNewBL} \]
\[ S_{DGAMod} \triangleq S_{DMod} + f(Maturity)S_{DGAmountModBL} \]
\[ S_{DGAmUmod} \triangleq S_{DUmod} + f(Maturity)S_{DGAmountUmodBL} \]

where

\[ f(Maturity) \equiv \text{Maturity adjustment factor function} \]
\[ S_{DGAmountNewBL} \equiv \text{Baseline (SDLCBegin) growth amount of New DSLOC expressed as a distribution of outcomes with associated attainment probability} \]
\[ S_{DGAmountModBL} \equiv \text{Baseline (SDLCBegin) growth amount of Modified DSLOC expressed as a distribution of outcomes with associated attainment probability} \]
\[ S_{DGAmountUmodBL} \equiv \text{Baseline (SDLCBegin) growth amount of Unmodified DSLOC expressed as a distribution of outcomes with associated attainment probability} \]

**Technical Baseline Estimated DSLOC Amounts**

The DEGM8 accepts, as input, Technical Baseline Estimate (TBE) amounts for New, Modified, and Unmodified DSLOC \((S_{DNew}, S_{DMod}, \text{and} \ S_{DUmod})\). These TBEs, often called point estimates, are rendered at various times during the program; typically by the program’s technical team based on some combination of engineering analysis, relevant past program experience, and expert judgment. These estimates represent the technical team’s best guess as to what the final outcome New, Modified, and Unmodified DSLOC values will be when the system is delivered and accepted. Note that these estimates are subject to error (size estimation uncertainty) caused by but not limited to (Ross, 2005):

- Technical team inability to accurately and precisely characterize the requirements as they are known in terms of New, Modified, and Unmodified DSLOC,
- DSLOC definition understanding, differences, and ambiguities,
- Automated code counting inconsistencies and discrepancies,
- Human error in completing the SRDR form,
- Human programmatic bias.
Maturity-Adjusted Growth Amounts

Each maturity-adjusted DSLOC growth amount is calculated as its particular baseline DSLOC growth amount (SDLCBegin to SwAccept) scaled by (multiplied by) its associated maturity adjustment factor. Note that the amount by which DSLOC grows (or shrinks) is subject to error caused by but not limited to (Ross, 2005):

- The customer doesn’t know what he/she wants,
- The customer doesn’t understand the problem,
- The mission has changed,
- The regulations that govern how this software should behave have changed,
- The vendor added a few extra features that he/she thought the customer would like,
- The project got behind schedule resulting in some requirements being dropped or postponed,
- The vendor finished early so the customer and/or the vendor thought up a few things to add.

Analysis of the preceding list suggests the following possible organization of issues that influence software size growth:

- Operational environment volatility
- Essence (requirements) volatility
- Essence understanding (requirements completeness and correctness)
- Essence versus implementation correspondence

Maturity Adjustment Factor

The maturity adjustment factor component of the DEGM8 represents the portion (percentage) of the baseline (SDLCBegin) DSLOC growth amount that remains to be experienced by the program as a function of where in the SDLC the TBE amounts for New, Modified, and Unmodified DSLOC are rendered by the program’s technical team. Obviously, if the estimates are rendered at SDLCBegin (i.e., are based on the knowledge that one would expect the program’s technical team to have at SDLCBegin), then the maturity adjustment factor should be 100% (i.e., all of the baseline DSLOC growth amount remains to be experienced). The DEGM8 assumes that, as the program progresses through the SDLC, a sequence of progressively-more-informed DSLOC estimates will be performed by the program’s technical team and that the corresponding maturity adjustment factors associated with each of these estimates will monotonically decrease and
approach 0% at delivery and acceptance of the software (i.e., a point in the SDLC where virtually all of the growth has been realized).

**Normalized Estimate Maturity**

In addition to the TBE amounts for New, Modified, and Unmodified DSLOC, the DEGM8 also accepts, as input, normalized estimate *Maturity*. Normalized estimate *Maturity* is assumed to be the *earned* percentage of the program’s *actual* (final) SDLCBegin to SwAccept duration that has elapsed at the time the estimate is rendered by the program’s technical team. Since the actual SDLCBegin to SwAccept duration cannot be known until SwAccept has occurred, the DEGM8 provides a surrogate quantification of earned duration percentages associated with certain key milestones. This default quantification is contained in Table 1 below. For example, if the program’s technical team performs an updated estimate of total New, Modified, and Unmodified DSLOC at Software Preliminary Design Review (SwPDR) then *Maturity* = 40% . Note that if the DEGM8 were to be used at SwPDR using the TBE DSLOC values rendered at SDLCBegin then *Maturity* = 0%; i.e., *Maturity* represents the maturity of the estimate. Estimate *Maturity* does not represent the maturity of the program or the maturity of the system under development. Note that there is some significant variability in when SRDR DD Form 2630-2’s are submitted, which contributes to the variability in DEGM8 estimates.

**Table 1** Default normalized estimate maturity scale

<table>
<thead>
<tr>
<th>SDLC Milestones</th>
<th>Software Development Life Cycle (SDLC)</th>
<th>Maturity of DSLOC Estimate</th>
<th>Life Cycle Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>DoD 5000 Milestones</td>
<td>Technology Development Phase (TD)</td>
<td>Technology Development Phase (TD)</td>
<td>Sys Rqmts and Software Feasibility</td>
</tr>
<tr>
<td>SDLC Milestones</td>
<td>Engineering &amp; Manufacturing Development Phase (EMD)</td>
<td>Engineering &amp; Manufacturing Development Phase (EMD)</td>
<td>Sys Design and Software Plans &amp; Requirements</td>
</tr>
<tr>
<td>Maturity of DSLOC Estimate</td>
<td>Technology Development Phase (TD)</td>
<td>Engineering &amp; Manufacturing Development Phase (EMD)</td>
<td>Product Design</td>
</tr>
<tr>
<td>SDLC Milestones</td>
<td>Technology Development Phase (TD)</td>
<td>Engineering &amp; Manufacturing Development Phase (EMD)</td>
<td>Detailed Design</td>
</tr>
<tr>
<td>Maturity of DSLOC Estimate</td>
<td>Technology Development Phase (TD)</td>
<td>Engineering &amp; Manufacturing Development Phase (EMD)</td>
<td>Code and Unit Test (CUT)</td>
</tr>
<tr>
<td>SDLC Milestones</td>
<td>Technology Development Phase (TD)</td>
<td>Engineering &amp; Manufacturing Development Phase (EMD)</td>
<td>Integration and Test (IT)</td>
</tr>
<tr>
<td>Maturity of DSLOC Estimate</td>
<td>Technology Development Phase (TD)</td>
<td>Engineering &amp; Manufacturing Development Phase (EMD)</td>
<td>Implementation</td>
</tr>
</tbody>
</table>

SDLCs come in all manner of activities, sequencing, and detail; the SDLC depicted in Table 1 is but one example. Customized versions of the milestone-maturity mapping can be developed based on alternative SDLCs given some knowledge of where in the SDLC’s elapsed duration certain selected milestones occur; however, milestones associated with the endpoints (*Maturity* = 0% and *Maturity* = 100%) must correspond to the two points in the SDLC where, respectively, the first SRDR DD Form 2630-2 and the final DD Form 2630-3 are submitted.

Note that it is best to use completed (earned) milestones to determine *Maturity* rather than just using the percentage of the currently-scheduled duration that has
elapsed. This is because the currently-scheduled duration is an estimate (i.e., it is never necessarily the final actual duration until the program is complete) and subject to change as the program progresses.

**Growth Decay**

The DEGM8 assumes some normalized maturity adjustment factor function $f$ of normalized $\text{Maturity}$ scaled such that $f(Maturity) \in [0,1]$, where $f(Maturity \mid Maturity = 0)$ represents the maximum (full scale) adjustment factor value (i.e., all the DSLOC growth has yet to be realized), and hypothesizes that $f(Maturity)$ decreases (decays) at a rate proportional to its value (i.e., unrealized growth tends to decay (decrease) faster during the early stages of an SDLC when experience is low and tends to decay slower during the later stages of an SDLC when experience is high). We model this hypothetical behavior mathematically as

$$\frac{d f(Maturity)}{d(Maturity)} \propto -f(Maturity)$$

(2)

where

$\propto \equiv \text{Proportionality relation; the left expression is directly proportional to the right expression}$

$\text{Decay} \equiv \text{Constant of proportionality}$

Solving the ordinary differential Equation (2) yields

$$\frac{d f(Maturity)}{f(Maturity)} = -(\text{Decay})d(Maturity)$$

$$\Rightarrow \int \frac{d f(Maturity)}{f(Maturity)} = \int -(\text{Decay})d(Maturity)$$

(3)

$$\Rightarrow \ln(f(Maturity)) = -(\text{Decay})(Maturity) + c$$

$$\therefore f(Maturity) = e^{-\text{(Decay)}(Maturity)}e^c$$

Since we have already posited the constraint $f(Maturity \mid Maturity = 0) = 1$ we can solve Equation (3) for the constant of integration $c$ as
Substituting the equivalent of \( c \) in Equation (4) for \( c \) in Equation (3) yields the maturity adjustment factor \( f(Maturity) \) portion of the DEGM8 equations in Figure 1 as

\[
f(0) = e^{-(Decay)(0)} e^{c} = 1 \rightarrow e^{c} = 1. \quad =c = 0
\]

In order to render the DEGM8 equations in Figure 1 useful in a particular estimating situation, we need to assume some value (or possibly some distribution) for the decay constant \( Decay \). Two methods for accomplishing this are:

1. to perform a regression analysis of relevant historical data to determine an expected value \( Decay \) or distribution \( Decay \)

2. to assume a value for \( Decay \) consistent with the slope of Boehm’s (1981 pp. 310-311) Cone of Uncertainty. Given the dearth of granular, periodic, and relevant historical DSLOC estimate data available to the authors at the time of this study, the latter method is used as the DEGM8’s default position. It is accomplished by scaling the top half of Boehm’s Cone of Uncertainty to be consistent with the SDLC milestones and percentages in Table 1 and fitting an exponential curve over the scaled result. As shown in Figure 2 below, a near-perfect fit can be achieved with \( Decay = 3.466 \). Note that the DEGM8 assumes only the curvature (shape) of Boehm’s Cone of Uncertainty and not its implied scaling (growth percentages).
Figure 2  Curve fit of the top half of the Boehm Cone of Uncertainty – is near perfect when the fit function is assumed to be \( y = 2e^{-3.466x} \).

Baseline DSLOC Growth Amounts

Without making any assumptions other than that the initial TBEs for New, Modified, and Unmodified DSLOC are related in some way to their respective final outcome values, we posit the following three definitions:

\[
\begin{align*}
S_{DGAmountNewBL} &\equiv f_{GN} \left( S_{DNew} \right) - S_{DNew} \\
S_{DGAmountModBL} &\equiv f_{GM} \left( S_{DMod} \right) - S_{DMod} \\
S_{DGAmountUmodBL} &\equiv f_{GU} \left( S_{DUmod} \right) - S_{DUmod}
\end{align*}
\]  

(6)

where

\[
\begin{align*}
S_{DGAmountNewBL} &\equiv \text{Baseline (SDLCEBegin) growth amount portion of New DSLOC expressed as a distribution of outcomes with associated attainment probability} \\
S_{DGAmountModBL} &\equiv \text{Baseline (SDLCEBegin) growth amount portion of Modified DSLOC expressed as a distribution of outcomes with associated attainment probability} \\
S_{DGAmountUmodBL} &\equiv \text{Baseline (SDLCEBegin) growth amount portion of Unmodified DSLOC expressed as a distribution of outcomes with associated attainment probability} \\
S_{DNew} &\equiv \text{Technical Baseline Estimate (TBE) of New DSLOC}
\end{align*}
\]
Baseline DSLOC Estimate Growth Relationships

For the sake of economy, we will show only the mathematical derivation for specifying $f_{GN}(S_{DNew})$ above and assume that the same process can be similarly applied to the specification of $f_{GM}(S_{DMod})$ and $f_{GU}(S_{DuUmod})$. We make certain assumptions about the nature of these relationships, the details of which are described as this section of the paper progresses. Making these assumptions and performing the linear algebra implied by these assumptions, we can elaborate $f_{GN}(S_{DNew})$ as follows:

\[
S_{DGANewBL} \triangleq \bar{b}_{GN} \varepsilon_{GN} S_{DNew}^{a_{GN}}
\]

where

$\propto$ \quad \equiv \text{Proportionality operator; the left operand is directly proportional to the right operand.}

$\triangleq$ \quad \equiv \text{Estimator equality symbol; the right expression estimates the left expression}

$\varepsilon_{GN}$ \quad \equiv \text{Baseline (SDLCBegin) New DSLOC growth error factor parameter distribution of outcomes with associated attainment probability}

$a_{GN}$ \quad \equiv \text{Exponent value parameter for the baseline New DSLOC growth estimating relationship that is calculated by the regression process; this exponent models the nonlinearity (economy or diseconomy of scale) present in the relationship between the initial DSLOC estimate and the final DSLOC actual; the baseline instance of the DEGM8 has a specific value for $a_{GN}$}

$\bar{b}_{GN}$ \quad \equiv \text{Geometric mean (log space arithmetic mean) scale factor parameter for the New DSLOC growth estimating relationship that is calculated by the regression process}

So how do we specify values for the exponent value parameter $a_{GN}$, the scale factor parameter $\bar{b}_{GN}$, and the error factor parameter distribution $\varepsilon_{GN}$ in Equation (7)? Ob-
viously any data-driven methodology must start with some data. In this case we require lists of completed Software Item (SI) initially-estimated New, Modified, and Unmodified DSLOC values with lists of corresponding final-actual New, Modified, and Unmodified DSLOC values. These lists we collectively refer to as an historical data set.

**DEGM8 Relevant Data**

**SRDR Data Filtering**

A primary objective when stratifying (filtering) historical data sets is to maximize the similarity between included observations while at the same time maximizing the number of observations included in the set. This objective, while often difficult to achieve, is nonetheless intended to both reduce the amount of variability and to increase the statistical significance of the relationships that we derive from the data. The baseline (default) instance of the DEGM8 equation parameter values for New, Modified, and Unmodified DSLOC is based on a subset of Software Resources Data Report (SRDR) data collected and archived by the U.S. Department of Defense’s Defense Cost and Resource Center (DCARC); this subset containing what we heretofore refer to as relevant data and satisfying the following filter criteria:

- **SerialNo2015**: >0 – observation must be included in the 2015 instance of the SRDR database (note that there are 8 observations in the 2011 SRDR database that do not appear in the 2015 SRDR database and have been assigned a Serial-No2015 value of 0)
- **Report**: 2630-3 – the source document for the observation must be a DD Form 2630-3, which documents the final actuals of a software development project (i.e., not estimated values); note that each DD Form 2630-3 observation in the SRDR database includes its initial-estimate DSLOC values from its associated DD Form 2630-2 form when such form exists
- **SI**: TRUE – the observation must represent a Computer Software Configuration Item (CSCI)-like Software Item (SI) (i.e., not a collection, summary, or roll-up of multiple CSCIs)
- **Nonphysical**: TRUE – the observation’s DSLOC values must not be measured in units of straight physical lines of code (i.e., they must be measured in logical lines of code (language statements) or non-comment physical lines of code); note that a lack of consistent code counting standards and techniques is a significant source of error in DEGM8 estimates
- **GFValid**: TRUE – the observation must contain values for:
  - **New_DSLOC_GF** – New DSLOC implied growth factor; ratio of final actual New DSLOC (SRDR DD Form 2630-3) to initial estimated New DSLOC (SRDR DD Form 2630-2)
Modified_DSLOC_GF – Modified DSLOC implied growth factor; ratio of final actual Modified DSLOC (SRDR DD Form 2630-3) to initial estimated Modified DSLOC (SRDR DD Form 2630-2)

Unmodified_DSLOC_GF – Unmodified DSLOC implied growth factor; ratio of final actual Unmodified DSLOC (SRDR DD Form 2630-3) to initial estimated Unmodified DSLOC (SRDR DD Form 2630-2)

that are all inside three geometric standard deviations from their respective population (entire database) geometric mean (see Table 2 below).

Table 2 Statistical outlier filtering comparison; regression JCDER349 with 3 geometric standard deviation statistical outlier filtering was chosen as the basis for the DEGM8

<table>
<thead>
<tr>
<th>Statistical Outlier Filtering:</th>
<th>New DSLOC</th>
<th>Modified DSLOC</th>
<th>Unmodified DSLOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Data Points (observations):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>None</td>
<td>3 GeoSigma</td>
<td>2 GeoSigma</td>
</tr>
<tr>
<td>Geometric (log space) mean of b:</td>
<td>0.7947</td>
<td>1.2084</td>
<td>1.1137</td>
</tr>
<tr>
<td>Arithmetic (unit space) mean of b:</td>
<td>3.4927</td>
<td>1.7360</td>
<td>1.4867</td>
</tr>
<tr>
<td>Standard deviation b:</td>
<td>1.8493</td>
<td>1.3418</td>
<td>1.4790</td>
</tr>
<tr>
<td>Coefficient of Variation (CV) b:</td>
<td>5.49</td>
<td>1.07</td>
<td>0.90</td>
</tr>
<tr>
<td>Arithmetic (unit space) mean of c:</td>
<td>1.9368</td>
<td>1.3665</td>
<td>1.2819</td>
</tr>
<tr>
<td>Standard deviation of c:</td>
<td>3.2795</td>
<td>1.2238</td>
<td>0.9590</td>
</tr>
<tr>
<td>Coefficient of Variation (CV) of c:</td>
<td>1.69</td>
<td>0.90</td>
<td>0.75</td>
</tr>
<tr>
<td>Mean Magnitude of the Relative Error:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Implied Growth Factor at data set arithmetic mean baseline DSLOC:</td>
<td>99% at 74,958</td>
<td>53% at 59,443</td>
<td>49% at 60,213</td>
</tr>
<tr>
<td>Implied Growth Factor at data set geometric mean baseline DSLOC:</td>
<td>80% at 25,635</td>
<td>50% at 23,035</td>
<td>45% at 23,672</td>
</tr>
</tbody>
</table>

See Appendix B for details and characteristics of the resulting relevant data set. Regarding the last filter criterion in the list above (GFValid), the geometric mean and the geometric standard deviation of a data set measure or metric are equivalent to the arithmetic mean and arithmetic standard deviation of that same measure or metric taken in log space. We choose the geometric mean and geometric standard deviation because they are more-suitable statistics for this historical data, the distributions of which have significant right skew. This is generally the case with software development essential measures historical data (Ross, 2008). The higher suitability comes from the fact that these geometric statistics provide an outlier determination that is more equitable to both high and low side outliers and thus tends to have less of an influence on the central tendency of the remaining observations.

The authors recognize that choosing to perform statistical filtering of outlier observations, as is done with GFValid above, is subject to some criticism; however, the statistics from the resulting data set show significant reduction in the Coefficients of Variation (CV) to values the authors consider somewhat more reasonable at the risk of...
possibly being unjustifiably more optimistic. The authors acknowledge the point of view
that suggests a no-statistical-filtering strategy might have been more appropriate since
it might have more-completely captured the inherent uncertainty. Whether or not this
filtering is valid depends on the amount of uncertainty that is due to the ineffectiveness
(or lack thereof) of the SRDR data collection, validation, and certification process versus
the amount of uncertainty inherent in the SDLC process associated with that SRDR.
Experience with the SRDR database has led the authors to be confident in the
assumption that the former dominates the latter and are therefore using the statistical
filtering as a surrogate for better validation and certification of the SRDR data. With
regard to the specific statistical filtering that was used for the default instance of the
DEGM8, we chose three geometric standard deviations as the filtering interval over two
geometric standard deviations because the differences between the geometric means
and CVs of the two alternatives does not, we believe, justify the increased degree of
filtering implied by the two standard deviations alternative.

New, Modified, and Unmodified DSLOC Data Sets

Let’s assume we have an historical data set containing ordered lists of the relevant
measures that represent the initial-estimated values and the final-actual values for each
of New, Modified, and Unmodified DSLOC for a population of \( N \) completed SIs.

In the case of New DSLOC, we map each of the New initial-estimated DSLOC values
list \( S_{DNEst} \) and the New final-actual DSLOC values list \( S_{DNAct} \) to a different dimension
of 2-dimensional (\( \mathbb{R}^2 \)) space: \( S_{DNEst} \) is represented by displacement along the \( \hat{e}_1 \)-axis (x-axis) and \( S_{DNAct} \) is represented by displacement along the \( \hat{e}_2 \)-axis (y-axis).\(^8\) We
organize this historical data set of points as the matrix \( P \).

\[
\begin{bmatrix}
P & S_{DNEst1} & S_{DNAct1} \\
S_{DNEst2} & S_{DNAct2} \\
\vdots & \vdots \\
S_{DNEstN} & S_{DNActN} 
\end{bmatrix}
\]  

Choosing the Functional Form of the Model

The notions of lines and planes in analytic geometry and linear algebra are linear; however, we cannot claim that estimated and actual size can be modeled as a linear combi-
nation; in other words we cannot claim that the relationship between \( S_{DNEst} \) and \( S_{DNAct} \)
is additive.

\[
f(S_{DNEst}, S_{DNAct}) \propto c_1 S_{DNEst} + c_2 S_{DNAct} + c_3
\]
where \( c_1 \) and \( c_2 \) are scale factor parameters and \( c_3 \) is an offset parameter. Experience with software development project historical data has shown that the relationships between essential measures tend to be \textit{nonlinear} (in this case multiplicative; i.e., the \textit{amount} of change in the output is \textit{not} proportional to the \textit{amount} change in the input which implies the existence of an economy or diseconomy of scale) (Ross, 2008); i.e.,

\[
f(S_{DNEst}, S_{DNAc}) \rightarrow S_{DNEst}^{c_1} S_{DNAc}^{c_2} c_3 \tag{10}
\]

where \( c_1 \) and \( c_2 \) are exponent parameters and \( c_3 \) is a scale factor parameter.

\textit{Log Transformation Makes the Problem Linear}

One method for applying linear algebra to nonlinear problems is called \textit{log transformation}. If we apply the log transformation function \( g \) to the multiplicative function \( f \) above; i.e., if we take the natural logarithm of both sides of Equation (10) we get

\[
g(f) \equiv \ln(f(S_{DNEst}, S_{DNAc})) = \ln(S_{DNEst}^{c_1} S_{DNAc}^{c_2} c_3) \tag{11}
\]

which is a linear combination (additive). We refer to the function \( f \) above as the function in \textit{unit space} and the log-transformed function \( g \) as the function in \textit{log space} or \textit{fit space}. The idea is to transform the given problem to log space (i.e., to transform the given problem into something that looks linear), apply appropriate linear algebra to the log-transformed problem, and then transform the results back to unit space. Transforming the log-space result of the linear algebra back to unit space involves exponentiating the result; i.e., raising the Euler number \( e \) to the power of the result.

Log transformation provides the added benefit of making the scaling of each dimension somewhat (but not perfectly) \textit{commensurable} and \textit{scale invariant}. In the case we are describing here, scale commensurability and invariance are guaranteed by virtue of the fact that both dimensions are scaled in the same units of measure (DSLOC). We mention this because commensurable and invariant scaling are a prerequisite for Orthogonal Distance Regression, the regression method we will be using later in this section.

If we log transform our historical data set \( P \) in Equation (8) we get \( \ln(P) \).
Best Fit Line Through a Log-transformed Data Set

Now that we have the historical data organized, the next step in instantiating the exponent parameter value, the scale factor parameter value, and the error factor distribution is to find the best fit line through our $N$-element log-transformed data set of points $\ln(P)$.

Using the linear algebra definition of the parametric system of equations form of a line we can describe our desired best fit line as

$$L_{\text{Best Fit}} \equiv \begin{cases} P_{\text{Best Fit}} S_{\text{DNEst}} = P'_{\text{Best Fit}} S_{\text{DNEst}} + t a_{\text{DNEst}} \\ P_{\text{Best Fit}} S_{\text{DNAc t}} = P'_{\text{Best Fit}} S_{\text{DNAc t}} + t a_{\text{DNAc t}} \end{cases}$$

(13)

where

$$\begin{bmatrix} P_{\text{Best Fit}} S_{\text{DNEst}}' \\ P_{\text{Best Fit}} S_{\text{DNAc t}}' \end{bmatrix} \equiv \text{any point on } L_{\text{Best Fit}}$$

$$\begin{bmatrix} P'_{\text{Best Fit}} S_{\text{DNEst}}' \\ P'_{\text{Best Fit}} S_{\text{DNAc t}}' \end{bmatrix} \equiv \text{some known point on } L_{\text{Best Fit}}$$

$$\begin{bmatrix} a_{\text{DNEst}} \\ a_{\text{DNAc t}} \end{bmatrix} \equiv \text{a direction vector of } L_{\text{Best Fit}}$$

$$t \in \mathbb{R} \equiv \text{the scaling parameter of } L_{\text{Best Fit}}$$

Orthogonal Distance Regression

At this point we choose to define best fit line as the Orthogonal Distance Regression (ODR) line; i.e., the line that results from minimizing the sum of the squared orthogonal (shortest) distances from the log-transformed data set points $\ln(P)$ to an ODR best fit line $L_{\text{ODR}}$ (see Appendix D); i.e., $L_{\text{Best Fit}} \equiv L_{\text{ODR}}$. Note that the application of ODR is a special case of Total Least Squares Regression.10
Specifying the ODR best fit line $L_{ODR}$ involves finding some point on $L_{ODR}$ and finding a direction vector $a$ that is parallel to $L_{ODR}$. We prove, in Appendix D, that the data set centroid of a data set always lies on its ODR best fit line; therefore, $L_{ODR}$ can be specified in terms of the direction vector $a$ and the centroid $C_{ln(P)}$ of our log-transformed data set as the system of parametric equations shown below in Equations (14) where $t$ is a scaling parameter.

$$L_{ODR} = \begin{cases}
    P_{ODR_{SDNest}} = C_{ln(P)_{SDNest}} + ta_{SDNest} \\
    P_{ODR_{SDNAct}} = C_{ln(P)_{SDNAct}} + ta_{SDNAct}
\end{cases} \tag{14}$$

Since we can compute the centroid $\left( C_{ln(P)_{SDNest}}, C_{ln(P)_{SDNAct}} \right)$ from our log-transformed data set $ln(P)$ using the equations

$$C_{ln(P)_{SDNest}} = \text{average}(S_{DNEst}) = \frac{1}{N} \sum_{i=1}^{N} ln(S_{DNEst_i})$$

$$C_{ln(P)_{SDNAct}} = \text{average}(S_{DNAct}) = \frac{1}{N} \sum_{i=1}^{N} ln(S_{DNAct_i}) \tag{15}$$

we need only find values for the coordinates of the direction vector $a$ in order to fully specify $L_{ODR}$.

**Singular Value Decomposition**

We now use the Singular Value Decomposition (SVD) to solve for the components of the $L_{ODR}$ direction vector $a$. See Appendix D for more detail about the SVD. In order to use the SVD for our purpose, we must first center our matrix about the origin of $\mathbb{R}^2$. We do this by subtracting, from each data set coordinate, its corresponding centroid coordinate as is shown in Equation (16). By doing this we create matrix $M$, a $2 \times N$ matrix, the centroid of which lies at the origin of $\mathbb{R}^2$; i.e., at the point $(0,0)$.

$$M = \begin{bmatrix}
    ln(S_{DNEst_1}) - C_{ln(P)_{SDNest}} & ln(S_{DNAct_1}) - C_{ln(P)_{SDNAct}} \\
    ln(S_{DNEst_2}) - C_{ln(P)_{SDNest}} & ln(S_{DNAct_2}) - C_{ln(P)_{SDNAct}} \\
    \vdots & \vdots \\
    ln(S_{DNEst_N}) - C_{ln(P)_{SDNest}} & ln(S_{DNAct_N}) - C_{ln(P)_{SDNAct}}
\end{bmatrix} \tag{16}$$
We next perform an SVD of matrix $M$. \textit{SVD}($M$) is a special factorization of $M$ such that

\begin{equation}
\text{SVD}(M) \equiv \{U, \Sigma, V^T\} \quad M = U\Sigma V^T
\end{equation}

where

$M$ \quad \equiv \quad \text{Given origin-centered data set matrix } (N \times 2)$

$U$ \quad \equiv \quad \text{Orthogonal matrix } (N \times 2); \text{ not used as part of the ODR process}$

$\Sigma$ \quad \equiv \quad \text{(Greek sigma, not to be confused with summation) Square diagonal matrix } (2 \times 2); \text{ contains the singular values of } M$

$V$ \quad \equiv \quad \text{Orthogonal matrix } (2 \times 2); \text{ the transpose of this matrix } V^T \text{ (as shown below) contains the singular vectors of } M \text{ organized as column vectors}$

\begin{equation}
M = U \begin{bmatrix}
\Sigma_{1,1} & 0 \\
0 & \Sigma_{2,2}
\end{bmatrix}
\begin{bmatrix}
V_{1,1} & V_{1,2} \\
V_{2,1} & V_{2,2}
\end{bmatrix}^T
= U \begin{bmatrix}
\Sigma_{\text{max}} & 0 \\
0 & \Sigma_{2,2}
\end{bmatrix}
\begin{bmatrix}
a_{\text{S-DNEst}} & V_{1,2} \\
a_{\text{S-DNAct}} & V_{2,2}
\end{bmatrix}^T
\end{equation}

The ODR best fit line’s direction vector $\mathbf{a}$ is the singular vector of $M$ that corresponds to the largest singular value of $M$; in this case either $\Sigma_{1,1}$ or $\Sigma_{2,2}$. A feature of the algorithm we are using to implement the SVD is that it returns $\Sigma$ and $V$ such that their contained values / vectors are sorted in descending singular value order from left to right. This implies the vector we are looking for is always in the leftmost column of $V$ as shown in Equation (18).

\textit{Instantiating the ODR Best-Fit Line}

We now have values for the components of direction vector $\mathbf{a}$ from Equation (18) and the components of the log-transformed data set centroid $C_{\text{ln}(p)}$, which can be used to specify the $L_{\text{ODR}}$ system of Equations (14). Since the ODR best fit line in log space represents our desired estimating relationship, it follows that any log-transformed single-point estimate of growth-adjusted New DSLOC $S_{\text{DGANewBL}}$, as represented by the point $P = (\text{ln}(S_{\text{DNew}}), \text{ln}(S_{\text{DGANewBL}}))$, must necessarily lie on $L_{\text{ODR}}$. Substituting the coordi-
nates of this single-point estimate \( P \) into Equations (14) and then solving each of the equations for the scalar \( t \) gives us

\[
L_{ODR} \equiv \begin{cases} 
  t = \frac{\ln(S_{DNew}) - C_{ln(p)}_{SDNEst}}{a_{SDNEst}} \\
  t = \frac{\ln(S_{DGANewBL}) - C_{ln(p)}_{SDNAcl}}{a_{SDNAcl}}
\end{cases}
\]  
\((19)\)

The right side expression of each equation in the \( L_{ODR} \) system of Equations (19) is equal to \( t \); therefore, we can set these expressions equal to each other.

\[
\frac{\ln(S_{DNew}) - C_{ln(p)}_{SDNEst}}{a_{SDNEst}} = \frac{\ln(S_{DGANewBL}) - C_{ln(p)}_{SDNAcl}}{a_{SDNAcl}}
\]  \((20)\)

Solving Equation (20) for \( \ln(S_{DGANewBL}) \), the estimated value we are looking for, yields

\[
\ln(S_{DGANewBL}) = \frac{a_{SDNAcl}}{a_{SDNEst}} \ln(S_{DNew}) C_{ln(p)}_{SDNEst} + C_{ln(p)}_{SDNAcl}
\]  \((21)\)

**Transforming the ODR Best-Fit Line to Unit Space**

We now transform our log space Equation (21) to unit space by exponentiating both sides of the equation; i.e., taking each of the two equivalent expressions as a power of \( e \) such that \( \exp(x) \leftrightarrow e^x \). This leads to the following algebraic manipulation and simplification:
\[
\exp\left(\ln\left(S_{DGANewBL}\right)\right) = \exp\left(\frac{a_{SDAnct}}{a_{SDNEst}} \left(\ln\left(S_{DNew}\right) - C_{ln(p)}_{SDNEst}\right)\right) \left(\exp\left(C_{ln(p)}_{SDNAct}\right)\right)
\]

\[
\rightarrow S_{DGANewBL} = \exp\left(\frac{a_{SDAnct}}{a_{SDNEst}} \ln\left(S_{DNew}\right) - C_{ln(p)}_{SDNEst}\right) \left(\exp\left(C_{ln(p)}_{SDNAct}\right)\right)
\]

\[
\rightarrow S_{DGANewBL} = \exp\left(\frac{a_{SDAnct}}{a_{SDNEst}} \ln\left(S_{DNew}\right)\right) \left(\exp\left(C_{ln(p)}_{SDNAct}\right)\right)
\]

\[
\therefore S_{DGANewBL} = \exp\left(\frac{a_{SDAnct}}{a_{SDNEst}} C_{ln(p)}_{SDNEst}\right) \left(\exp\left(C_{ln(p)}_{SDNAct}\right)\right)
\]

Letting \( a_{GN} = \frac{a_{SDAnct}}{a_{SDNEst}} \) and \( b_{GN} = \frac{\exp\left(C_{ln(p)}_{SDNAct}\right)}{\exp\left(\frac{a_{SDAnct}}{a_{SDNEst}} C_{ln(p)}_{SDNEst}\right)} \), and substituting \( a_{GN} \) and \( b_{GN} \) in Equation (22) with their equivalent expressions above gives us

\[
S_{DGANewBL} = b_{GN} S_{DNew}^{a_{GN}}
\]

The application of the SVD gives us specific values for \( a_{SDAnct} \) and \( a_{SDNEst} \), and, therefore, a specific value for \( a_{GN} \). However we, as yet, have no specific value for \( b_{GN} \). We can remedy this by first instantiating Equation (23) with the appropriate data set lists \( S_{DNEst} \) and \( S_{DAnct} \) of observations as defined in Equation (8) and then solve for the list of corresponding scale factor parameter values \( b_{GN} \) to get the list \( b_{GN} \).

\[
\left(\begin{array}{c}
b_{GN_i} \\
S_{DNEst_i}^{a_{GN}}
\end{array}\right)_{i=1}^{N} = b_{GN} \left(\frac{\sum_{i=1}^{N} S_{DAnct_i}^{a_{GN}}}{\sum_{i=1}^{N} S_{DNEst_i}^{a_{GN}}}\right)
\]

We then use an appropriate central tendency value of the list \( b_{GN} \) to estimate \( b_{GN} \). Since \( b_{GN} \) exists in unit space but is based on ODR performed in log space, we choose to use the arithmetic mean of \( b_{GN} \) in log space (geometric mean of \( b_{GN} \) in unit space) as
the estimator of $b_{GN}$. Since we can now compute a value $\tilde{b}_{GN}$ for the geometric mean of the list $b_{GN}$,

$$\tilde{b}_{GN} \equiv \text{GeoMean}(b_{GN}) = \exp \left( \frac{1}{n} \sum_{i=1}^{N} \ln(b_{GN_i}) \right)$$  \hspace{1cm} (25)$$

we instantiate Equation (23) with the value $\tilde{b}_{GN}$ to get

$$S_{DGANewBL} \triangleq \tilde{b}_{GN} S_{DNew} a_{GN}$$  \hspace{1cm} (26)$$

**Error and Uncertainty**

*Orthogonal Distance Represents the Error*

Notice that Equation (26) is a single-point (non-probabilistic) version of the baseline New DSLOC estimate growth Equation (7). It estimates specific baseline growth-adjusted New DSLOC as a function of specific single values of $a_{GN}$, $\tilde{b}_{GN}$, and TBE New DSLOC $S_{DNew}$. We refer to Equation (26) as being expressed in *estimator form*, recognized by the fact that baseline growth-adjusted DSLOC $S_{DGANewBL}$ is being estimated by the expression to the right of the “$\triangleq$” symbol.

A reality of estimation is the existence of *error* in both the independent and dependent variables of an estimating relationship. There exist numerous technological, programmatic, personnel, and modelling factors that together determine the magnitude and direction of this error (Ross, 2003). Earlier in the paper we listed several causes for this error specific to software growth. When using ODR, the resultant orthogonal error vector $n_{e_{GN_i}}$ for a given observation data point $\ln(P_i) = [\ln(S_{DNew_i}), \ln(S_{DNAct_i})]$ is the vector $Q_i \ln(P_i)$ where point $Q_i$ is on the best fit line $L_{ODR}$ and where $Q_i \ln(P_i)$ is orthogonal (normal) to $L_{ODR}$. The point $Q_i$ necessarily represents the point on $L_{ODR}$ that is closest to the point $\ln(P_i)$. In our case, each error vector $n_{e_{GN_i}}$ has one component vector in each of the two dimensions of our New DSLOC data subset. One of these component vectors $e_{GNest_i}$ is associated with the independent variable $\ln(S_{DNew_i})$ and is parallel to the $\hat{e}_1$-axis (x-axis) and the other component vector $e_{GNAct_i}$ is associated with the dependent variable $\ln(S_{DNAct_i})$ and is parallel to the $\hat{e}_2$-axis (y-axis). The geometry of the resultant error and its components is illustrated in Figure 3 below.
Analytic geometry provides that, for a line given by the equation \( ax + by + c = 0 \) where \( a \), \( b \), and \( c \) are real constants with \( a \) and \( b \) not both zero, the point \( (x, y) \) on this line that is closest to some point \( (x_0, y_0) \) has the coordinates

\[
x = \frac{b(x_0 - ay_0) - ac}{a^2 + b^2} \quad \text{and} \quad y = \frac{a(-bx_0 + ay_0) - bc}{a^2 + b^2}
\]

Equation (27)

We can rewrite Equation (21) to match the \( ax + by + c = 0 \) form as

\[
\ln\left( S_{DGA_{NewBL}} \right) = \frac{a_{SDNA_{Act}}}{a_{SDNEst}} \left( \ln\left( S_{DNew} \right) + C_{ln(p)_{SDNEst}}^+ \right) C_{ln(p)_{SDNA_{Act}}}
\]

\[
\therefore \frac{a_{SDNA_{Act}}}{a_{SDNEst}} \ln\left( S_{DNew} \right) - \ln\left( S_{DGA_{NewBL}} \right) + C_{ln(p)_{SDNA_{Act}}} - \frac{a_{SDNA_{Act}}}{a_{SDNEst}} C_{ln(p)_{SDNEst}} = 0
\]

which implies

\[
a = \frac{a_{SDNA_{Act}}}{a_{SDNEst}} = a_{GN} \quad \text{(previously defined)}
\]
\[ b \equiv -1 \]
\[ c \equiv C_{ln(p)}^{SDNAct} - \frac{a_{SDNAct} C_{ln(p)}^{SDNEst}}{a_{SDNEst}} \]

We let the point \((x_0, y_0)\) represent a log-transformed observation data point
\[ ln(P_i) = (ln(S_{DNEst i}), ln(S_{DNAct i})) \]; therefore,
\[ x_0 \equiv ln(S_{DNEst i}) \]
\[ y_0 \equiv ln(S_{DNAct i}) \]

Making the appropriate substitutions into Equations (27) and simplifying the result yields
\[ Q_i = (Q_{ix}, Q_{iy}) \]
\[ \left( \frac{(ln(S_{DNEst i}) + a_{GN} ln(S_{DNAct i})) - a_{GN} \left( C_{ln(p)}^{SDNAct} - a_{GN} C_{ln(p)}^{SDNEst} \right)}{a_{GN}^2 + 1} \right) \]
\[ \left( \frac{a_{GN} \left( ln(S_{DNEst i}) + a_{GN} ln(S_{DNAct i}) \right) + \left( C_{ln(p)}^{SDNAct} - a_{GN} C_{ln(p)}^{SDNEst} \right)}{a_{GN}^2 + 1} \right) \]
\[ (29) \]

We can now define each of the signed magnitudes of each error component as
\[ \varepsilon_{SDNEst i} \equiv ln(S_{DNEst i}) - Q_{ix} \] (30)
and
\[ \varepsilon_{SDNAct i} \equiv ln(S_{DNAct i}) - Q_{iy} \] (31)

using Equation (29) to instantiate each of \( Q_{ix} \) and \( Q_{iy} \).

We can instantiate Equations (30) and (31) with the appropriate lists \( S_{DNEst} \), \( S_{DNAct} \), \( Q_x \), and \( Q_y \), to get corresponding error lists \( \varepsilon_{SDNEst} \) and \( \varepsilon_{SDNAct} \) as
\[ \left( \varepsilon_{SDNEst i} \equiv ln(S_{DNEst i}) - Q_{ix} \right)_{i=1}^N \rightarrow \varepsilon_{SDNEst} = ln(S_{DNEst}) \cdot Q_x \] (32)
and
\[ \left( \varepsilon_{SDNAct i} \equiv ln(S_{DNAct i}) - Q_{iy} \right)_{i=1}^N \rightarrow \varepsilon_{SDNAct} = ln(S_{DNAct}) \cdot Q_y \] (33)
One advantage of using ODR instead of an Ordinary Least Squares (OLS)-based regression method is this ability to isolate the error in each of the estimating relationship’s variables rather than using just the dependent variable’s residual value to represent the error.

Because we have developed our estimating relationship using ODR in log space; the error, while additive in log space; must be transformed to unit space and treated as multiplicative. We can rewrite Equation (26) to account for this multiplicative error in each variable (Equations (32) and (33)) as

\[
S_{DGANewBL} \exp(\varepsilon_{SDNAct}) \triangleq \tilde{b}_{GN} \left( S_{DNew} \exp(\varepsilon_{SDNEst}) \right)^{a_{GN}}
\]

where

\( \varepsilon_{SDNEst} \) = Log space error component list associated with the list \( \ln(S_{DNEst}) \)

\( \varepsilon_{SDNAct} \) = Log space error component list associated with the list \( \ln(S_{DAct}) \)

Letting

\( \varepsilon_{GN} \equiv \frac{\exp(\varepsilon_{SDNEst})^{a_{GN}}}{\exp(\varepsilon_{SDNAct})} \)  \hspace{1cm} (35)

and then substituting \( \varepsilon_{GN} \) for its equivalent into Equation (34) gives us

\[
S_{DGANewBL} \triangleq \tilde{b}_{GN} \varepsilon_{GN} S_{DNew}^{a_{GN}}
\]

If we were to assume that the historical values in the two corresponding \( \ln(S_{DNEst}) \) and \( \ln(S_{DAct}) \) lists of log-transformed relevant measures are perfectly correlated; i.e., each and every historical observation in the data set \( \ln(P_i) = [\ln(S_{DNEst_i}) \ln(S_{DAct_i})] \) lies on the ODR best-fit line \( L_{ODR} \), then every element of the error factor parameter list \( \varepsilon_{GN} \) in Equation (36) would have the same value (specifically a value of 1); therefore, the list \( \varepsilon_{GN} \) would be unnecessary and the list \( S_{DGANewBL} \) would simply be the single
value $S_{DGANewBLS}$. The estimator form of the baseline New DSLOC estimate growth equation sans $\epsilon_{GN}$ would be sufficient to do perfectly-certain (i.e., deterministic) estimating with respect to some given value of $S_{DNew}$. This scenario, however, is not realistic since, for each observation, error is almost always present in each of the relevant measures. We will henceforth refer to the net effect of this error as uncertainty since its presence is what makes estimation uncertain; i.e., stochastic or probabilistic and not deterministic or a sure thing. Since Equation (36) contains only three parameters, $a_{GN}$, $b_{GN}$, and $\epsilon_{GN}$; and since $a_{GN}$ and $b_{GN}$ are constants of the regression, it follows that the range and distribution of the values in the $\epsilon_{GN}$ list can be used to model the uncertainty in Equation (36). Because the list $\epsilon_{GN}$ is being used to model a range of possible outcomes, we must indicate Equation (36) as a probabilistic estimating relationship by replacing the calibration form list variable $\epsilon_{GN}$ with its corresponding estimator form random (distribution) variable $\epsilon_{GN}$. Since the variable $S_{DGANewBLS}$, represented in Equation (36) as the list of estimates $S_{DGANewBLS}$, is now being estimated as a function of $\epsilon_{GN}$, it follows that $S_{DGANewBLS}$ is now the random variable $S_{DGANewBLS}$. Note that $S_{DNew}$ remains a single-value variable since we have specified that this be given as a single-value TBE. Therefore,

$$S_{DGANewBLS} \triangleq b_{GN} \epsilon_{GN} S_{DNew} a_{GN}$$

(37)

which is identical to Equation (7).

Custom CDFs Approximate Random Variable Distributions

One of the features of TRI’s ACE tool (as is the case with other similar statistical tools) is that it supports the notion of defining random variables in terms of what ACE refers to as a custom Cumulative Distribution Function (CDF) (see Endnote 1). We wish to approximate $\epsilon_{GN}$ with a custom CDF that is based on the list $\epsilon_{GN}$; therefore, $\epsilon_{GN} \equiv \text{customCDF}(\epsilon_{GN})$ where $\text{customCDF}(\epsilon_{GN})$ represents a process that returns an $N \times 2$ matrix $C$, where $N$ is the number of elements in $\epsilon_{GN}$, such that

$$C_{1,2}, C_{2,2} \ldots C_{N,2} \equiv \text{ascending_sort}(\epsilon_{GN})$$

$$C_{1,1}, C_{2,1} \ldots C_{N,1} \equiv \text{percentile_rank}(C_{1,2}, C_{2,2} \ldots C_{N,2})$$

(38)

The result is a matrix $C$ where the second column contains an ascending-sorted list of the values from $\epsilon_{GN}$ and where the first column contains the percentile rank of its associated value in the second column. This matrix $C$ can be copied into ACE as the defini-
tion of a custom CDF that can be used to approximate the probability distribution associated with $\varepsilon_{GN}$.

**Equations for Baseline New, Modified, and Unmodified DSLOC**

If we apply the same overall baseline DSLOC estimate growth relationship derivation, that we used for New DSLOC above, to each of Modified and Unmodified DSLOC we get all three equations we are looking for:

$$
S_{DGANewBL} \triangleq \tilde{b}_{GN} \varepsilon_{GN} S_{DNew} a_{GN} \\
S_{DGAModBL} \triangleq \tilde{b}_{GM} \varepsilon_{GM} S_{DMod} a_{GM} \\
S_{DGAUmodBL} \triangleq \tilde{b}_{GU} \varepsilon_{GU} S_{DUmod} a_{GU}
$$

(39)

Custom CDFs for each of $\varepsilon_{GN}$, $\varepsilon_{GM}$, and $\varepsilon_{GU}$ that are specific to the default instance of the DEGM8 can be found in Appendix A.

**Software Item Normalization Factor**

The observations in the relevant data set being used as the basis for the default instance of the DEGM8 are the result of overall SRDR database filtering that includes only those observations that represent Computer Software Configuration Item (CSCI)-like Software Items (SIs). In cases where the DEGM8 is being used to estimate growth-adjusted DSLOC for SIs other than CSCI-like SIs, a normalization of $S_{DNew}$, $S_{DMod}$, and $S_{DUmod}$ is necessary to make the scaling of these three values consistent with the historical data in order to prevent disproportionate application of the economy or diseconomy of scale effects resulting from the exponents $a_{GN}$, $a_{GM}$, and $a_{GU}$. We therefore enhance Equations (39) to incorporate this scaling normalization as

$$
S_{DGANewBL} \triangleq \tilde{b}_{GN} \varepsilon_{GN} \frac{S_{DNew}}{K_N} a_{GN} K_N \\
S_{DGAModBL} \triangleq \tilde{b}_{GM} \varepsilon_{GM} \frac{S_{DMod}}{K_M} a_{GM} K_M \\
S_{DGAUmodBL} \triangleq \tilde{b}_{GU} \varepsilon_{GU} \frac{S_{DUmod}}{K_U} a_{GU} K_U
$$

(40)

where

$K_N, K_M, K_U$ \quad \equiv \quad \text{Software Item (SI) to Computer Software Configuration Item (CSCI) normalization factors for New, Modified, and Un-}
modified DSLOC such that each equals the number of CSCIs represented by the SI; for example, if the SI being estimated is one of four equal-size components of a CSCI, then each of $K_N$, $K_M$, and $K_U$ would equal 0.25. Likewise, if the SI being estimated is a collection of four CSCIs, then each of $K_N$, $K_M$, and $K_U$ would equal 4.

**Converting Growth-Adjusted Totals to Growth Amounts**

Notice that, as written, Equations (40) estimate baseline growth-adjusted *total amounts* of DSLOC; however, we will next be applying a maturity adjustment but only wish to apply that adjustment to the *growth amount* portion of the total and not to the whole growth-adjusted total. Therefore, we rewrite Equations (40) to include subtracting the given TBE DSLOC amounts from their associated equations in order to isolate the portions of growth-adjusted DSLOC that represent the amount of DSLOC growth.

\[
\begin{align*}
S_{DGAmountNewBL} &\triangleq \tilde{b}_{GN} \epsilon_{GN} \frac{S_{DNew}}{K_N} a_{GN} K_N - S_{DNew} \\
S_{DGAmountModBL} &\triangleq \tilde{b}_{GM} \epsilon_{GM} \frac{S_{DMod}}{K_M} a_{GM} K_M - S_{DMod} \\
S_{DGAmountUmodBL} &\triangleq \tilde{b}_{GU} \epsilon_{GU} \frac{S_{DUmod}}{K_U} a_{GU} K_U - S_{DUmod}
\end{align*}
\]

(41)

**Assembling the Model Components**

**Maturity-Adjusted Growth Amount**

We begin with the baseline (SDLCBegin) DSLOC estimate growth amount distributions of Equations (41), which have been developed from historical data and which model the amount of uncertainty that exists about the TBEs of New, Modified, and Unmodified DSLOC. We assume that these estimates are done at the beginning of the SDLC; i.e., $Maturity = 0$, consistent with the SDLCs from which the historical data were collected. Suppose these baseline distributions are represented as CDFs; i.e., mappings of growth amount values to probability of attainment percentages. We would like to model what happens to the uncertainty modeled by these distributions as activities in the SDLC progress to completion and TBEs are updated to reflect the evolving knowledge. We have already hypothesized that uncertainty decays over time and have developed a model for this decay in Equation (5). Since the maturity adjustment factor function of Equation (5) is normalized (i.e., yields factors that are percentages of full scale), we can use this factor...
to scale our baseline estimated growth amounts of New, Modified, and Unmodified DSLOC to reflect the estimate maturity of subsequent TBEs.

Equations (41) represent the baseline estimated growth amounts of New, Modified, and Unmodified DSLOC. We wish to adjust these amounts as a function of estimate Maturity (i.e., when in the SDLC the estimate is rendered). We accomplish this by multiplying these amounts by the maturity adjustment factor function Equation (5) to yield the expressions

\[
\begin{align*}
&f(Maturity)S_{DGAmountNewBL} \\
&f(Maturity)S_{DGAmountModBL} \\
&f(Maturity)S_{DGAmountUmodBL}
\end{align*}
\] (42)

We can expand the three Expressions (42) by substituting each factor with its equivalent in Equations (5) and (41) respectively, which gives us the maturity-adjusted growth amounts

\[
\begin{align*}
e^{-\text{(Decay)(Maturity)}}b_{GN}e_{GN} & \left( \frac{S_{DNew}}{K_N} \right)^{a_{GN}} K_N - S_{DNew} \\
e^{-\text{(Decay)(Maturity)}}b_{GM}e_{GM} & \left( \frac{S_{DMod}}{K_M} \right)^{a_{GM}} K_M - S_{DMod} \\
e^{-\text{(Decay)(Maturity)}}b_{GU}e_{GU} & \left( \frac{S_{DUmod}}{K_U} \right)^{a_{GU}} K_U - S_{DUmod}
\end{align*}
\] (43)

Notice that the estimate maturity factor portion of Expressions (43) serves to decay (exponentially decrease) the estimated DSLOC growth amount distributions. The practical effect of applying this decay is time-progressive compression of the effective DSLOC growth factor distributions about the TBE position approaching no uncertainty at SDLC completion (SWAccept, Maturity = 100%).

As stated earlier, in order to render Expressions (43) useful in a particular estimating situation, we need to assume some value (or distribution) for the decay constant Decay in Expressions (43); either by assuming Decay = 3.466 (Boehm’s Cone of Uncertainty) or by analyzing relevant historical data to model decay as a single value Decay or as a distribution Decay.
Total Growth-Adjusted DSLOC Estimate Amounts

We now simply add the TBEs for New, Modified, and Unmodified DSLOC to each of the Expressions (43) to produce the final form DEGM8 equations:

\[
S_{DGANew} \triangleq S_{DNew} + e^{-\text{Decay}\times\text{Maturity}} \left[ \tilde{b}_{GN} \bar{e}_{GN} \left( \frac{S_{DNew}}{K_N} \right)^{a_{GN}} K_N - S_{DNew} \right]
\]

\[
S_{DGAMod} \triangleq S_{DMod} + e^{-\text{Decay}\times\text{Maturity}} \left[ \tilde{b}_{GM} \bar{e}_{GM} \left( \frac{S_{DMod}}{K_M} \right)^{a_{GM}} K_M - S_{DMod} \right]
\]

(44)

\[
S_{DGAUmod} \triangleq S_{DUmod} + e^{-\text{Decay}\times\text{Maturity}} \left[ \tilde{b}_{GU} \bar{e}_{GU} \left( \frac{S_{DUmod}}{K_U} \right)^{a_{GU}} K_U - S_{DUmod} \right]
\]

Note that these equations match those of Figure 1.

Since each of the right-side expressions of Equations (44) represent total growth-adjusted DSLOC amounts, we can divide each expression by its associated TBE DSLOC amount to create associated expressions that represent implied growth factors which we can use as a metric to describe the effective growth being applied by the model to the TBEs of New, Modified, and Unmodified DSLOC:

\[
\begin{align*}
S_{DNew} + e^{-\text{Decay}\times\text{Maturity}} \left[ \tilde{b}_{GN} \bar{e}_{GN} \left( \frac{S_{DNew}}{K_N} \right)^{a_{GN}} K_N - S_{DNew} \right] & \Bigg/ S_{DNew} \\
S_{DMod} + e^{-\text{Decay}\times\text{Maturity}} \left[ \tilde{b}_{GM} \bar{e}_{GM} \left( \frac{S_{DMod}}{K_M} \right)^{a_{GM}} K_M - S_{DMod} \right] & \Bigg/ S_{DMod} \\
S_{DUmod} + e^{-\text{Decay}\times\text{Maturity}} \left[ \tilde{b}_{GU} \bar{e}_{GU} \left( \frac{S_{DUmod}}{K_U} \right)^{a_{GU}} K_U - S_{DUmod} \right] & \Bigg/ S_{DUmod}
\end{align*}
\]

(45)

Correlation is Necessary to Properly Add Random Variables

Equations (44) yield random variables for each of growth-adjusted New, Modified, and Unmodified DSLOC. It is likely, as part of an overall estimation process of which the DEGM8 is a part, that there will be a need to define some random variable as the sum of \(S_{DGANew}\), \(S_{DGAModified}\), and \(S_{DGAUmodified}\) (i.e., total DSLOC) or to define some random variable as the sum of random variables that are functions of \(S_{DGANew}\),
\( S_{DGAModified} \), and \( S_{DGAUnmodified} \). The process of summing random variables is called convolution\(^\text{13}\), the mathematics of which are beyond the scope of this paper. However, several commercially available software tools provide the capability to sum both independent (uncorrelated) and dependent (correlated) random variables. TRI’s ACEIT is one such software tool; convolution is performed using Monte Carlo simulation techniques and correlation is applied using an internal algorithm that requires providing pairwise correlation between the summands. The DEGM8 provides this pairwise correlation in the form of a correlation matrix that is developed as part of the overall historical data regression process. Table 3 shows the correlation matrix for the default instance of the DEGM8. The three values from this matrix necessary as inputs to ACEIT in order to properly convolve growth-adjusted New, Modified, and Unmodified DSLOC are:

\[
\text{correl}(b_{GN}, b_{GM}) = 2.56966251 \times 10^{-03} \\
\text{correl}(b_{GN}, b_{GU}) = 3.02466226 \times 10^{-01} \\
\text{correl}(b_{GM}, b_{GU}) = 7.46696118 \times 10^{-02}
\]

(46)

Table 3  Correlation matrix from JCDER349 regression (DEGM8 default instance)

<table>
<thead>
<tr>
<th>Correlation Matrix</th>
<th>New DSLOC</th>
<th>Modified DSLOC</th>
<th>Unmodified DSLOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>New DSLOC</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modified DSLOC</td>
<td>2.569663E-03</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Unmodified DSLOC</td>
<td>3.024662E-01</td>
<td>7.466961E-02</td>
<td>1</td>
</tr>
</tbody>
</table>

4. DEGM8 EXAMPLE APPLICATION

In this section of the paper we describe an example of how we can use the default version of the DEGM8 to adjust TBEs of New, Modified, and Unmodified DSLOC to develop growth-adjusted probabilistic estimates of what we expect these values to be when the software is done and accepted. Appendix E contains a description and tables that show how this example can be implemented in TRI’s ACEIT software tool.

Ground Rules and Assumptions for the Example

For this example, we are given the task of providing a growth-adjusted probabilistic size estimate for the Navigation (NAV) Function Computer Software Configuration Item (CSCI) that is part of an unmanned satellite ground control system. The following list represents the Ground Rules and Assumptions (GRAs) that describe the scenario we will be using:
• The given TBEs for New, Modified, and Unmodified software size are 25,000 DSLOC, 50,000 DSLOC, and 100,000 DSLOC respectively.
• We assume no redelivered code from a previous increment, iteration, or release of the NAV function.
• Since the NAV function is a CSCI; i.e., it is of a size and scope that is normally associated with a single observation in the SRDR database, we assume $K[N] = 1$, $K[M] = 1$, and $K[U] = 1$; i.e., normalization of the TBEs to the historical data is unnecessary.
• We assume that the TBEs for New, Modified, and Unmodified DSLOC are rendered at the completion of Software Requirements Review (SwRR); therefore, $Maturity \equiv 20\%$ per Table 1.
• We assume $Decay \equiv 3.466$ based on Boehm’s (1981 pp. 310-311) Cone of Uncertainty as is specified for the default instance of the DEGM8.
• We assume the DEGM8 default values for the exponent parameters associated with each of New, Modified, and Unmodified DSLOC as $a[GN] = 1.021$, $a[GM] = 0.913$, and $a[GU] = 1.044$ per the second page of Appendix B.
• We assume the DEGM8 default values for the geometric means of the scale factor parameters associated with each of New, Modified, and Unmodified DSLOC as $\bar{b}[GN] = 1.208$, $\bar{b}[GM] = 2.651$, and $\bar{b}[GU] = 0.6199$ per the second page of Appendix B.
• We assume the DEGM8 default correlation values shown in Equations (46).

**Instantiating the DEGM Equations**

Given the GRAs listed above, we can instantiate the DEGM equations in Figure 1 as
\[
S_{DGANew} \triangleq S_{DNew} + e^{-(\text{Decay})(\text{Maturity})} \left( \tilde{b}_{GN} \varepsilon_{GN} \left( \frac{S_{DNew}}{K_N} \right)^{a_{GN}} K_N - S_{DNew} \right)
\]
\[
S_{DGAMod} \triangleq S_{DMod} + e^{-(\text{Decay})(\text{Maturity})} \left( \tilde{b}_{GM} \varepsilon_{GM} \left( \frac{S_{DMod}}{K_M} \right)^{a_{GM}} K_M - S_{DMod} \right)
\]
\[
S_{DGAUmod} \triangleq S_{DUmod} + e^{-(\text{Decay})(\text{Maturity})} \left( \tilde{b}_{GU} \varepsilon_{GU} \left( \frac{S_{DUmod}}{1} \right)^{a_{GU}} K_U - S_{DUmod} \right)
\]

\[
S_{DGANew} \triangleq 25,000 + e^{-(3.466)(20\%)} \left( 1.208 \varepsilon_{GN} \left( \frac{25,000}{1} \right)^{1.021} (1) - 25,000 \right)
\]
\[
S_{DGAMod} \triangleq 50,000 + e^{-(3.466)(20\%)} \left( 2.651 \varepsilon_{GM} \left( \frac{50,000}{1} \right)^{0.913} (1) - 50,000 \right)
\]
\[
S_{DGAUmod} \triangleq 100,000 + e^{-(3.466)(20\%)} \left( 0.6199 \varepsilon_{GU} \left( \frac{100,000}{1} \right)^{1.044} (1) - 100,000 \right)
\]

Figure 4, Figure 5, and Figure 6 below illustrate the behaviors of the resulting growth-adjusted New DSLOC, Modified DSLOC, and Unmodified DSLOC estimate distributions as described in Equations (47) for the given TBEs and estimate Maturity.
Figure 4: Example growth-adjusted New DSLOC distribution vs. estimate maturity

Figure 5: Example growth-adjusted Modified DSLOC distribution vs. estimate maturity
Implied Growth Factors as the Estimates Mature

We can instantiate the implied growth factor Expressions (45) with values from our example’s GRAs to get
We can use Expressions (48) to solve for the New, Modified, and Unmodified DSLOC growth factors implied by the DEGM8 as a function of increasing estimate \( Maturity \). Figure 7, Figure 8, and Figure 9 below illustrate this behavior of over the range of possible estimate \( Maturity \) values \( Maturity \in [0\%, 100\%] \).
Figure 7 New DSLOC implied growth-factor decay; the implied growth factor for a TBE of 25,000 New DSLOC at SwRR (\textit{Maturity} = 20\%) is 1.52 (52\%) at the arithmetic mean (61\% percentile) and 1.25 (25\%) at the geometric mean (41\% percentile).

Figure 8 Modified DSLOC implied growth-factor decay; the implied growth factor for a TBE of 50,000 Modified DSLOC at SwRR (\textit{Maturity} = 20\%) is 1.28 (28\%) at the arithmetic mean (75\% percentile) and 1.02 (2\%) at the geometric mean (43\% percentile).
Figure 9  Unmodified DSLOC implied growth-factor decay; the implied growth factor for a TBE of 100,000 Unmodified DSLOC at SwRR \((Maturity = 20\%)\) is 1.11 (11\%) at the arithmetic mean (77th percentile) and 1.01 (1\%) at the geometric mean (46th percentile).

5. A SPECIAL INSTANCE OF DEGM8 FOR ESTIMATING UNMANNED SPACECRAFT FLIGHT SOFTWARE

A particular challenge for USAF Space and Missile Systems Center (SMC) is estimating the size of unmanned Space Vehicle (SV) embedded software (bus software and payload software). One thing that makes this a challenge is the lack of readily-available relevant historical data. The SRDR database, as of this writing, contains very few observations of this kind of software. Apparently, Government Contractors have typically not been required to furnish SV software measurement and metric data to the Government in the form of submitted SRDRs.

SMC with assistance from Tecolote Research, Inc. has collected some relevant historical DSLOC estimates and final-actual DSLOC results for several space programs in order to improve on existing growth, cost, and duration estimates for unmanned SV software; however, the estimates were performed at various points in the SDLC and most of the programs they represent did not perform DSLOC estimates at SDLCBegin. As such, this data is insufficient to develop Growth Estimating Relationships (GERs) using the approach that was taken for the default instance of the DEGM8. The remainder of this section describes a new and unique variant approach and resulting GERs that can be used for SV programs to perform growth-adjusted New, Modified, and Unmodified DSLOC estimates as a function of TBE New, Modified, and Unmodified...
DSLOC size; and estimate maturity. This model variant is henceforth referred to as DEGM8SV and is based on relevant data collected from programs where there exists a final-actual DSLOC result and at least one DSLOC estimate along with its corresponding estimate maturity.

Once again, for the sake of economy, we focus on New DSLOC and assume that the same process will hold for Modified and Unmodified DSLOC.

**Identify the Desired Functional Form**

We start by recognizing the fact that we have three unique data fields for each historical observation: $S_{DN Est}$, $S_{DN Act}$, and $Maturity_{GN}$. Since we wish to maturity-adjust only the growth portion of a TBE and not the entire estimate, we have created a new implied growth factor field $S_{DN GF} \equiv \frac{S_{DN Act}}{S_{DN Est}}$ that represents the amount of growth yet to be realized. It is to this field that we can multiply the decaying estimate $Maturity$ adjustment factor $e^{-(Decay_{GN})(Maturity_{GN})}$ that we developed earlier in Equation (5).

Using the above observations and assumptions, we propose the following multiplicative combination with which we will derive and specify our GERs for New, Modified, and Unmodified DSLOC.

$$
S_{DN Est} e^c_1 S_{DN GF} e^c_2 e^{c_3 (Maturity_{GN})} c_4 \equiv S_{DN GF} \frac{S_{DN Act}}{S_{DN Est}}
$$

**Normalize the Measures to Ensure Scale Invariance and Commensurability**

In order to ensure scale invariance and measurement commensurability we sigma-normalize each of the historical data $N$-element lists $S_{DN Est}$, $S_{DN GF}$, and $Maturity_{GN}$.
\[
\begin{align*}
S'_{\text{DNEst}} & \equiv \frac{S_{\text{DNEst}}}{\sigma_{\text{DNEst}}} = \frac{S_{\text{DNEst}}}{\text{stdev}(S_{\text{DNEst}})} \\
S'_{\text{DNGF}} & \equiv \frac{\ln(S_{\text{DNGF}})}{\sigma_{\ln(S_{\text{DNGF}})}} = \frac{\ln(S_{\text{DNGF}})}{\text{stdev}(\ln(S_{\text{DNGF}}))} \\
\text{Maturity'}_{\text{GN}} & \equiv \frac{\text{Maturity}_{\text{GN}}}{\sigma_{\text{Maturity}_{\text{GN}}}} = \frac{\text{Maturity}_{\text{GN}}}{\text{stdev}(\text{Maturity}_{\text{GN}})}
\end{align*}
\]

\[
\begin{align*}
\left(1 - \frac{1}{N} \sum_{i=1}^{N} (S_{\text{DNEst},i} - \mu_{\text{DNEst}})^2\right)^{\frac{1}{2}} & \equiv \mu_{\text{DNEst}} = \frac{1}{N} \sum_{k=1}^{N} S_{\text{DNEst},k} \\
\left(1 - \frac{1}{N} \sum_{i=1}^{N} (S_{\text{DNGF},i} - \mu_{\text{DNGF}})^2\right)^{\frac{1}{2}} & \equiv \mu_{\text{DNGF}} = \frac{1}{N} \sum_{k=1}^{N} S_{\text{DNGF},k} \\
\left(1 - \frac{1}{N} \sum_{i=1}^{N} (\ln(\text{Maturity}_{\text{GN},i}) - \mu_{\text{Maturity}_{\text{GN}}})^2\right)^{\frac{1}{2}} & \equiv \mu_{\text{Maturity}_{\text{GN}}} = \frac{1}{N} \sum_{k=1}^{N} \text{Maturity}_{\text{GN},k}
\end{align*}
\]

Log Transform the Data to Make the Relationship Linear

\[
g(f) \equiv \ln\left(f\left(S'_{\text{DNEst}}, S'_{\text{DNGF}}, e^{(\text{Maturity'}_{\text{GN}})}\right)\right) = \ln\left(S'_{\text{DNEst}} e^{c_1}, S'_{\text{DNGF}} e^{c_2}, e^{c_3(\text{Maturity'}_{\text{GN}})} e^{c_4}\right)
\]

\[
\therefore g(f) = c_1 \ln(S'_{\text{DNEst}}) + c_2 \ln(S'_{\text{DNGF}}) + c_3(\text{Maturity'}_{\text{GN}}) + \ln(c_4)
\]

Organize the Historical Data as a Matrix

\[
\ln'(P) = \begin{bmatrix}
\ln(S'_{\text{DNEst},1}) & \ln(S'_{\text{DNGF},1}) & \text{Maturity'}_{\text{GN},1} \\
\ln(S'_{\text{DNEst},2}) & \ln(S'_{\text{DNGF},2}) & \text{Maturity'}_{\text{GN},2} \\
\vdots & \vdots & \vdots \\
\ln(S'_{\text{DNEst},N}) & \ln(S'_{\text{DNGF},N}) & \text{Maturity'}_{\text{GN},N}
\end{bmatrix}
\]
Define the ODR Best Fit Line

\[
L_{ODR} = \begin{cases} 
    P_{ODR_{DNEst}}' = P_{ODR_{DNEst}}' + ta_{DNEst} \\
    P_{ODR_{DNGF}}' = P_{ODR_{DNGF}}' + ta_{DNGF} \\
    P_{ODR_{Maturity_{GN}}} = P_{ODR_{Maturity_{GN}}} + ta_{Maturity_{GN}} 
\end{cases} \quad (53)
\]

Find the Data Set Centroid Point Coordinates

\[
\begin{aligned}
C_{ln(p)}_{S_{DNEst}} &= \text{average} \left( \ln \left( S'_{DNEst} \right) \right) = \frac{1}{N} \sum_{i=1}^{N} \ln \left( S'_{DNEst_i} \right), \\
C_{ln(p)}_{S_{DNGF}} &= \text{average} \left( \ln \left( S'_{DNGF} \right) \right) = \frac{1}{N} \sum_{i=1}^{N} \ln' \left( S'_{DNGF_i} \right), \\
C_{ln(p)}_{Maturity_{GN}} &= \text{average} \left( \text{Maturity}'_{GN} \right) = \frac{1}{N} \sum_{i=1}^{N} \text{Maturity}'_{GN_i}
\end{aligned} \quad (54)
\]

Instantiate the ODR Best Fit Line with the Data Set Centroid

\[
L_{ODR} = \begin{cases} 
    P_{ODR_{DNEst}} = C_{ln(p)}_{S_{DNEst}} + ta_{S_{DNEst}} \\
    P_{ODR_{DNGF}} = C_{ln(p)}_{S_{DNGF}} + ta_{S_{DNGF}} \\
    P_{ODR_{Maturity_{GN}}} = C_{p_{Maturity_{GN}}} + ta_{Maturity_{GN}} 
\end{cases} \quad (55)
\]

Center the Data Set Matrix about the Coordinate System Origin

\[
M = \begin{bmatrix}
\ln \left( S'_{DNEst_1} \right) - C_{ln(p)}_{S_{DNEst}} & \ln \left( S'_{DNEst_2} \right) - C_{ln(p)}_{S_{DNEst}} & \ldots & \ln \left( S'_{DNEst_N} \right) - C_{ln(p)}_{S_{DNEst}} \\
\ln \left( S'_{DNGF_1} \right) - C_{ln(p)}_{S_{DNGF}} & \ln \left( S'_{DNGF_2} \right) - C_{ln(p)}_{S_{DNGF}} & \ldots & \ln \left( S'_{DNGF_N} \right) - C_{ln(p)}_{S_{DNGF}} \\
\vdots & \vdots & \ddots & \vdots \\
\ln \left( S'_{Maturity_{GN_1}} \right) - C_{ln(p)}_{S_{Maturity_{GN}}} & \ln \left( S'_{Maturity_{GN_2}} \right) - C_{ln(p)}_{S_{Maturity_{GN}}} & \ldots & \ln \left( S'_{Maturity_{GN_N}} \right) - C_{ln(p)}_{S_{Maturity_{GN}}}
\end{bmatrix} \quad (56)
\]

Apply the SVD to the Centered Data Set Matrix

\[
SVD \left( M \right) \equiv \left\{ U, \Sigma, V^T \right\} \quad M = U \Sigma V^T \quad (57)
\]
\[ SVD(M) \equiv \{U, \Sigma, V^T\} \bigg| M = U \Sigma V^T \]

\[
M = U \begin{bmatrix}
\Sigma_{1,1} & 0 & 0 \\
0 & \Sigma_{2,2} & 0 \\
0 & 0 & \Sigma_{3,3}
\end{bmatrix}
\begin{bmatrix}
V_{1,1} & V_{1,2} & V_{1,3} \\
V_{2,1} & V_{2,2} & V_{2,3} \\
V_{3,1} & V_{3,2} & V_{3,3}
\end{bmatrix}^T
\]

\[
M = U \begin{bmatrix}
\Sigma_{\text{max}} & 0 & 0 \\
0 & \Sigma_{2,2} & 0 \\
0 & 0 & \Sigma_{3,3}
\end{bmatrix}
\begin{bmatrix}
a_{SDNEst} & V_{1,2} & V_{1,3} \\
a_{SDNGF} & V_{2,2} & V_{2,3} \\
a_{\text{Maturity}_GN} & V_{3,2} & V_{3,3}
\end{bmatrix}^T
\]

\[ (58) \]

### Define the Growth Estimating Relationship Hyperplane

We now introduce and define the notion of a growth estimating relationship (GER) hyperplane. We define a GER hyperplane \( H_{\text{GER}} \) as a 2-dimension flat subspace of \( \mathbb{R}^3 \) (a plane in this case) that is orthogonal to the ODR best fit line \( L_{\text{ODR}} \) and that passes through a point on \( L_{\text{ODR}} \) regarded as a fundamental solution \( P_{\text{GER}} \) on \( L_{\text{ODR}} \). The importance of a GER hyperplane is that it contains all the points in \( \mathbb{R}^3 \) that orthogonally project onto \( L_{\text{ODR}} \) at \( P_{\text{GER}} \); i.e., \( P_{\text{GER}} \) is the fundamental solution for every point on \( H_{\text{GER}} \). This implies that the GER hyperplane \( H_{\text{GER}} \) contains all the potential outcome coordinate combinations that are best estimated by the coordinate combination represented by point \( P_{\text{GER}} \).

The nature of our desired GER is such that we wish to structure the GER hyperplane equation so that the final-actual estimate of New DSLOC growth amount \( S_{\text{DNGAmt}} \) is dependent on the TBE of New DSLOC \( S_{\text{DNew}} \). Note that in estimating situations we are not trying to solve for \( S_{\text{DNew}} \) but rather are given \( S_{\text{DNew}} \) as a single value based on when in the SDLC the estimate is being rendered; i.e., its estimate \( \text{Maturity} \). We instantiate the variables in the general scalar equation form of a hyperplane

**Theorem: Scalar or Implicit Form Equation of a Hyperplane in m-Dimension Space**

\[
\mathbf{n} \cdot \mathbf{r} + d = 0
\]

\[
n_1 r_1 + n_2 r_2 + \ldots + n_m r_m + d = 0
\]

where

\[
d \equiv -\mathbf{n} \cdot \mathbf{r}'
\]

\[
= n_1 r_1' + n_2 r_2' + \ldots + n_m r_m'
\]
using the following observations and assumptions, the result being an equation describing $H_{GER}$.

- Since the best fit GER hyperplane is orthogonal to our ODR best fit line and since a hyperplane is partially described by a vector that is orthogonal (normal) to that hyperplane, we let the direction vector $a$ from the definition of our ODR best fit line $L_{ODR}$ also be the normal vector $n$ of our JCDER hyperplane $H_{JCDER}$. Therefore, $n \equiv a$.

- The GER hyperplane $H_{GER}$ must include the fundamental solution point $P_{GER}$. We therefore use $P_{GER}$ as our known point on $H_{GER}$, the position vector of which is represented in Equation (59) as $r'$. Since $P_{GER}$ must lie on our ODR best fit line $L_{ODR}$ we can instantiate Equations (55) as

$$
\begin{align*}
L_{ODR} &\equiv \begin{cases} 
    r'_1 = C_{ln(p)SDNEst} + t a_{SDNEst} \\
    r'_2 = C_{ln(p)SDNGF} + t a_{SDNGF} \\
    r'_3 = C_{p_{MaturityGN}} + t a_{MaturityGN}
\end{cases} \\
(60)
\end{align*}
$$

Since $P_{GER}$ must also be dependent on the given value for TBE New DSLOC it follows that the $P_{GER}$ position vector $ln(S'_{DNEst})$ component must be

$$
 r'_1 \equiv ln(S'_{DNew}) \\
(61)
$$

Substituting $r'_1$ in the first of Equations (60) with its equivalent in Equation (61) yields

$$
ln(S'_{DNew}) = C_{ln(p)SDNEst} + t a_{SDNEst} \\
(62)
$$

Solving for the parameter $t$ in Equation (62) we get

$$
t = \frac{ln(S'_{DNew}) - C_{ln(p)SDNEst}}{a_{SDNEst}} \\
(63)
$$

Substituting each instance of $t$ in Equations (60) with its equivalent in Equation (63) gives us a New DSLOC TBE-constrained specification of the fundamental solution point $P_{GER}$ and therefore the components of the GER hyperplane known point’s position vector $r'$. 

42
\[ r'_1 = \ln(S'_{D\text{New}}) \]
\[ r'_2 = C_{\ln(p)_{SDNGF}} + \frac{\ln(S'_{D\text{New}}) - C_{\ln(p)_{SDNEst}}}{a_{SDNEst}} a_{SDNGF} \]  \hspace{1cm} (64) 
\[ r'_3 = C_{p_{Maturity_{GN}}} + \frac{\ln(S'_{D\text{New}}) - C_{\ln(p)_{SDNEst}}}{a_{SDNEst}} a_{Maturity_{GN}} \]

- Any potential sigma-normalized log-transformed estimate \((\ln(S'_{D\text{New}}), \ln(S'_{DNGF}), Maturity'_{GN})\) must lie somewhere on our GER hyperplane.

Therefore, \(r_1 = \ln(S'_{D\text{New}}), r_2 = \ln(S'_{DNGF})\), and \(r_3 = Maturity'_{GN}\).

This gives us the GER hyperplane

\[ a_{SDNEst} \ln(S'_{D\text{New}}) + a_{SDNGF} \ln(S'_{DNGF}) + a_{Maturity_{GN}} Maturity'_{GN} + d = 0 \]  \hspace{1cm} (65) 

where

\[ d = \begin{pmatrix}
-a_{SDNEst} \ln(S'_{D\text{New}}) - \\
a_{SDNGF} \left( C_{ln(p)_{SDNGF}} \frac{\ln(S'_{D\text{New}}) - C_{\ln(p)_{SDNEst}}}{a_{SDNEst}} a_{SDNGF} \right) \\
a_{Maturity_{GN}} \left( C_{p_{Maturity_{GN}}} \frac{\ln(S'_{D\text{New}}) - C_{\ln(p)_{SDNEst}}}{a_{SDNEst}} a_{Maturity_{GN}} \right)
\end{pmatrix} \]  \hspace{1cm} (66) 

Next, we substitute \(d\) in Equation (65) with its equivalent in Equation (66) to get

\[ a_{SDNEst} \ln(S'_{D\text{New}}) + a_{SDNGF} \ln(S'_{DNGF}) + a_{Maturity_{GN}} Maturity'_{GN} + \\
-a_{SDNEst} \ln(S'_{D\text{New}}) - \\
a_{SDNGF} \left( C_{ln(p)_{SDNGF}} \frac{\ln(S'_{D\text{New}}) - C_{\ln(p)_{SDNEst}}}{a_{SDNEst}} a_{SDNGF} \right) - \\
a_{Maturity_{GN}} \left( C_{p_{Maturity_{GN}}} \frac{\ln(S'_{D\text{New}}) - C_{\ln(p)_{SDNEst}}}{a_{SDNEst}} a_{Maturity_{GN}} \right) = 0 \]  \hspace{1cm} (67) 

Algebraically rearranging the factors and terms of Equation (67) gives us
\[ a_{SDNGF} \ln(S_{SDNGF}') = \left( \frac{a_{SDNGF}^2 + a_{MaturityGN}}{a_{SDNEst}} \right) \ln(S_{DNew}') - \left( \frac{a_{SDNGF}^2 + a_{MaturityGN}}{a_{SDNEst}} \right) C_{\ln(p)SDNEst} + a_{MaturityGN} C_{pMaturityGN} \]

Transforming Equation (68) to unit space yields

\[ \exp\left( a_{SDNGF} \ln(S_{SDNGF}') \right) = \exp\left( \left( \frac{a_{SDNGF}^2 + a_{MaturityGN}}{a_{SDNEst}} \right) \ln(S_{DNew}') \right) \exp\left( -a_{MaturityGN} Maturity'_GN \right) \]

\[ \exp\left( \left( \frac{a_{SDNGF}^2 + a_{MaturityGN}}{a_{SDNEst}} \right) C_{\ln(p)SDNEst} + a_{MaturityGN} C_{pMaturityGN} \right) \]

\[ \therefore S_{DNew}' = \left( \frac{a_{SDNGF}^2 + a_{MaturityGN}}{a_{SDNEst}} \right) \exp\left( -\left( \frac{a_{MaturityGN}}{a_{SDNGF}} \right) Maturity'_GN \right) \left( \exp\left( \frac{a_{SDNGF}^2 + a_{MaturityGN}}{a_{SDNEst}} \right) C_{\ln(p)SDNEst} + a_{MaturityGN} C_{pMaturityGN} \right) \]

We next convert the normalized variables \( S_{DNew}' \), \( S_{SDNGF}' \), and \( Maturity'_GN \) back to their un-normalized state by multiplying each by their corresponding standard deviations from Definitions (50) to get

\[ S_{DNGF} = \left( \frac{a_{SDNGF}^2 + a_{MaturityGN}}{a_{SDNEst}} \right) \exp\left( -\left( \frac{a_{MaturityGN}}{a_{SDNGF}} \right) Maturity'_GN \right) \left( \exp\left( \frac{a_{SDNGF}^2 + a_{MaturityGN}}{a_{SDNEst}} \right) C_{\ln(p)SDNEst} + a_{MaturityGN} C_{pMaturityGN} \right) \]
Letting $a_{GN}^{-1} \equiv \frac{a_{SDNGF}^2 + a_{Maturity_{GN}}^2}{a_{SDNEst}}$, $Decay_{GN} \equiv \frac{a_{Maturity_{GN}}}{a_{SDNGF}}$, and

$$b_{GN} = \exp \left( \frac{\left( a_{SDNGF}^2 + a_{Maturity_{GN}}^2 \right)^2}{a_{SDNEst}} C_{ln(p)}^{SDNEst} \right)$$

gives us

$$S_{DNGF} = \frac{S_{DNEst} a_{GN}^{-1} \exp\left( -Decay_{GN} Maturity_{GN} \right)}{b_{GN}^{-1}}$$

$$\therefore S_{DNGF} = e^{-\left( Decay_{GN} \right) \left( Maturity_{GN} \right) b_{GN} S_{DNew} a_{GN}^{-1}}$$

Acknowledging $b_{GN}$ as a list $b_{GN}$ with geometric mean central tendency $\bar{b}_{GN}$

$$\left( \begin{array}{c} b_{GN_i} = \frac{S_{DNGF_i}}{e^{-\left( Decay_{GN} \right) \left( Maturity_{GN} \right) S_{DNEst} a_{GN}^{-1}}} \end{array} \right)_{i=1}^{N}$$

$$\rightarrow b_{GN} = \frac{S_{DNGF}}{e^{-\left( Decay_{GN} \right) \left( Maturity_{GN} \right) S_{DNEst} a_{GN}^{-1}}}$$

$$\bar{b}_{GN} = GeoMean (b_{GN}) = \exp \left( \frac{1}{N} \sum_{i=1}^{N} ln\left( b_{GN_i} \right) \right)$$

we instantiate Equation (71) with the value $\bar{b}_{GN}$ to get

$$S_{DNGF} = e^{-\left( Decay_{GN} \right) \left( Maturity_{GN} \right) \bar{b}_{GN} S_{DNew} a_{GN}^{-1}}$$

**Apply Error and Uncertainty Attributed to the GER**

Linear algebra provides that in $\mathbb{R}^m$ space any line can be described by the vector equation

$$\mathbf{r} = \mathbf{r}' + a \mathbf{t} \quad \mathbf{r}, \mathbf{r}', \mathbf{a} \in \mathbb{R}^m, \quad t \in \mathbb{R}$$

where $\mathbf{r}$ is a position vector representing any point on that line, $\mathbf{r}'$ is a position vector representing a known point on that line, $\mathbf{a}$ is a direction vector of that line, and $t$ is a scaling parameter such that for every specific unique value of $t$ there is a specific unique value of $\mathbf{r}$. We can instantiate Equation (75) with what we already know about $L_{ODR}$ and the data upon which it is based:
• We have already formatted and translated (offset) all of the sigma-normalized and log-transformed data points as the matrix \( M \) such that the data set centroid \( C_{ln(p)}^{SDN} \) of the translated data set has the position vector \([0 \ 0 \ 0]\); aka, the zero vector \(0\).

• We have already proven that a best fit ODR line passes through (contains) its data set centroid; therefore, since our data set is now centroid-translated, we can let \( r' \equiv [0 \ 0 \ 0] \).

• We have already determined a direction vector that is appropriate for this best fit ODR line through the sigma-normalized log-transformed centroid-translated data set when we earlier applied the SVD process; therefore,

\[
\begin{bmatrix}
  a_{SDNEst} \\
  a_{SDNGF} \\
  a_{Maturity_{GN}}
\end{bmatrix}
\equiv S_S Maturity_{DNEst} DNGF GN.
\]

Making the appropriate substitution for the known point position vector \( r' \equiv [0 \ 0 \ 0] \) yields

\[
\begin{align*}
  r &\equiv [0 \ 0 \ 0] \\
  &\equiv t a_{SDN} \\
  \rightarrow r &\equiv t a_{SDN} \\
  &\in \mathbb{R}^m \\
  t &\in \mathbb{R}
\end{align*}
\]

(76)

Linear algebra also gives us a theorem for projecting one vector \( v \) in \( \mathbb{R}^m \) space onto another vector \( u \) in the same vector space

\[
proj_u(v) = \left( \frac{u \cdot v}{u \cdot u} \right) u
\]

(77)

We can find the position vector \( r_i'' \) of the point \( P_i'' \) on \( L_{ODR}' \) that is closest to a sigma-normalized log-transformed centroid-translated historical observation \( ln''(P_i) \) by projecting the \( ln''(P_i) \) position vector \( r_i'' \) onto the position vector \( r \) of any point on \( L_{ODR}' \). We set this up by letting \( proj_u(v) \equiv r'', u \equiv r \), and \( v \equiv r'' \); and substituting these equivalencies into Theorem (77) to get

\[
r_i'' = \left( \frac{r \cdot r''}{r \cdot r} \right) r
\]

(78)

Substituting \( r \) in Equation (78) with its equivalent \( t a_{SDN} \) in Equation (76) gives us
The error vector in sigma-normalized log-transformed centroid-translated space $\epsilon_i'$ associated with a data point $\ln^*(P_i)$ is expressed as the difference $r''_i - r''$ where each coordinate of $\epsilon_i'$ represents its associated dimensional component of the data point’s error vector. Therefore

$$\epsilon_i' \equiv r''_i - r'' \rightarrow r'' \equiv r''_i - \epsilon_i'$$

Substituting $r''_i$ in Equation (79) with its equivalent in Equation (80) gives us

$$r''_i - \epsilon_i' = \left( \frac{a_{SDN} \cdot r''_i}{a_{SDN} \cdot a_{SDN}} \right) a$$

$$\rightarrow \epsilon_i' = r''_i - \left( \frac{a_{SDN} \cdot r''_i}{a_{SDN} \cdot a_{SDN}} \right) a_{SDN}$$

We can now define the signed magnitudes of each New DSLOC estimate error vector component as

$$\left[ \epsilon_{SDNEst_i}' \quad \epsilon_{SDNEst_i}' \quad \epsilon_{MaturityGN_i}' \right] \equiv \epsilon_i = r''_i - \left( \frac{a_{SDN} \cdot r''_i}{a_{SDN} \cdot a_{SDN}} \right) a_{SDN}$$

We next instantiate and evaluate Equation (82) with each of our three sigma-normalized, log-transformed, centroid-translated lists $\ln(S''_{DNEst_i})$, $\ln(S''_{DNGF})$, and $\text{Maturity}_{GN}'$ to produce the sigma-normalized, log-transformed error lists $\epsilon_{SDNEst}'$, $\epsilon_{DNGF}'$, and $\epsilon_{MaturityGN}'$ as
Note that we do not refer to the error component lists as being centroid-translated since error is a relative quantity with respect to a data set and its best fit ODR line; therefore, if the data set and hence its best fit ODR line are re-positioned within the space, the error component lists are not affected.

We next convert the sigma-normalized log-transformed error component lists $\varepsilon_{SDNEst}'$, $\varepsilon_{SDNGF}'$, and $\varepsilon_{MaturityGN}'$ to un-sigma-normalized unit space by first exponentiating each of $\varepsilon_{SDNEst}'$ and $\varepsilon_{SDNGF}'$ (NOT $\varepsilon_{MaturityGN}'$ since it remains exponentiated in our unit space equation) and then by multiplying all three of $\varepsilon_{SDNEst}'$, $\varepsilon_{SDNGF}'$, and $\varepsilon_{MaturityGN}'$ by the standard deviation value that corresponds to its respective dimension (see Definitions (50)) to get $\varepsilon_{SDNEst}$, $\varepsilon_{SDNGF}$, and $\varepsilon_{MaturityGN}$.

Because we have developed our estimating relationship using ODR in log space; the error represented by $\varepsilon_{SDNEst}$ and $\varepsilon_{SDNGF}$, while additive in log space; must be transformed to unit space and treated as multiplicative. We can rewrite Equation (74) to account for this multiplicative error in each variable as

$$S_{DNGF} \cdot (\varepsilon_{SDNGF}) = e^{-(Decay_{GN})(Maturity_{GN} + \varepsilon_{Maturity})} \cdot b_{GN} \cdot S_{DNew} \cdot (\varepsilon_{SDNEst}) \cdot a_{GN}^{-1}$$

$$\therefore S_{DNGF} = e^{-(Decay_{GN})(Maturity_{GN})} \cdot b_{GN} \cdot (e^{\varepsilon_{SDNEst}} \cdot a_{GN}^{-1}) \cdot (\varepsilon_{SDNGF}) \cdot (S_{DNEst} \cdot a_{GN}^{-1})$$

Letting
\[ \varepsilon_{GN} = \left( e^{-\frac{a_{GN}^{a_{GN}^{-1}}}{\text{Decay}_{GN}\text{Maturity}_{GN}}} \right) \varepsilon_{SDNEst} \]  
\[(85)\]

and then substituting \( \varepsilon_{GN} \) for its equivalent into Equation (84) gives us

\[ S_{DNGF} \triangleq e^{-\left(\text{Decay}_{GN}\left(\text{Maturity}_{GN}\right)\right)} \tilde{b}_{GN} \varepsilon_{GN} S_{DNew} a_{GN}^{-1} \]  
\[(86)\]

Substituting \( S_{DNGF} \) in Equation (86) with its equivalent defined in the constraint portion of Equation (49) yields

\[ \frac{S_{DGANew}}{S_{DNew}} \triangleq e^{-\left(\text{Decay}_{GN}\left(\text{Maturity}_{GN}\right)\right)} \tilde{b}_{GN} \varepsilon_{GN} S_{DNew} a_{GN}^{-1} \]  
\[ \therefore S_{DGANew} = e^{-\left(\text{Decay}_{GN}\left(\text{Maturity}_{GN}\right)\right)} \tilde{b}_{GN} \varepsilon_{GN} S_{DNew} a_{GN} \]  
\[(87)\]

Express the GER as a Random Variable Estimating Relationship

\[ S_{DGANew} \triangleq e^{-\left(\text{Decay}_{GN}\left(\text{Maturity}_{GN}\right)\right)} \tilde{b}_{GN} \varepsilon_{GN} S_{DNew} a_{GN} \]  
\[(88)\]

Add CSCI Normalization and Extend to All 3 Types of DSLOC

\[ S_{DGANew} \triangleq e^{-\left(\text{Decay}_{GN}\left(\text{Maturity}_{GN}\right)\right)} \tilde{b}_{GN} \varepsilon_{GN} \left( \frac{S_{DNew}}{K_N} \right)^{a_{GN}} K_N \]  
\[ S_{DGAMod} \triangleq e^{-\left(\text{Decay}_{GM}\left(\text{Maturity}_{GM}\right)\right)} \tilde{b}_{GN} \varepsilon_{GM} \left( \frac{S_{DMod}}{K_M} \right)^{a_{GM}} K_M \]  
\[ S_{DGAMod} \triangleq e^{-\left(\text{Decay}_{GU}\left(\text{Maturity}_{GU}\right)\right)} \tilde{b}_{GU} \varepsilon_{GU} \left( \frac{S_{DGUmod}}{K_U} \right)^{a_{GU}} K_U \]  
\[(89)\]

Results

Custom CDFs for each of \( \varepsilon_{GN} \), \( \varepsilon_{GM} \), and \( \varepsilon_{GU} \) that are specific to DEGM8SV can be found in Appendix A. Measurements, parameters, and statistics specific to DEGM8SV can be found in the tables contained in Appendix F.
6. SUMMARY AND CONCLUSIONS

A significant challenge that many cost analysts and project managers face is predicting by how much their estimates of software development cost and schedule will change over the life of the project. Examination of currently-accepted software cost, schedule, and defect estimation algorithms reveals a common acknowledgment that estimated software size is the single most influential independent variable (Ross, 2003) (Ross, 2005). Unfortunately, the most important business decisions about a software project are made at its beginning; the time when most estimating is done; and coincidently the time of minimum knowledge, maximum uncertainty, and hysterical optimism (Ross, 2005). This paper describes a model and methodology, DEGM8, that provides probabilistic growth adjustment to single-point TBEs of DSLOC, for New, Modified, and Unmodified software that is sensitive to the maturity of the estimates. The model is based on more and more-recent SRDR data collected by the DoD combined with a state-of-the-art data regression technique (Orthogonal Distance Regression using Singular Value Decomposition) that more-accurately models the data and its error (uncertainty).

It continues to be the authors’ collective opinion that the DEGM8, as described in this paper, represents a quantum improvement over the field of available software code growth methodologies. Specifically, among the advantages of this model over the Holchin (2003) and Jensen (2008) code growth matrices are the following:

- DEGM8 is based on DoD-collected SRDR data versus Holchin’s Delphi survey of experts approach and Jensen’s data from multiple proprietary sources.
- DEGM8 requires only one parameter, estimate Maturity, which is reasonably objective versus Holchin’s and Jensen’s rather subjective and vaguely-defined Complexity and Maturity parameters.
- DEGM8 equations are non-linear allowing for the existence of economies/diseconomies of scale. Holchin and Jensen model growth as a constant factor (as does DEGM7).
- DEGM8 equations include a composite error factor distribution (embodies uncertainty derived from the data) versus Holchin’s single-point growth-factor result. (Jensen uses the lognormal distribution to model uncertainty.)
- DEGM8 provides error factor distribution decay based on updated estimate Maturity versus Holchin’s single-point growth factor reduction based on updated Complexity and Maturity parameters. (Jensen defines Maturity in terms of defined program phases.)
- DEGM8 differentiates between New, Modified, and Unmodified DSLOC growth versus Holchin’s and Jensen’s one-growth-factor-fits-all approach.
DEGM8 provides correlation (a correlation matrix) between each of growth-adjusted New, Modified, and Unmodified DSLOC to support proper summation of their distributions. Neither Holchin nor Jensen provide this capability.

The DEGM8 represents a significant update and improvement to the DEGM7, which has been used as part of the basis for numerous USAF Program Objective Memoranda (POM), Program Office Estimates (POEs), Independent Cost Estimates (ICEs), and Service Cost Positions (SCPs). Planned enhancements to this model include:

- Updating the model to incorporate new data from the 2017 update of the SRDR database;
- Rerunning the data analysis using specific stratifications of the SRDR database in order to create growth models that are unique to specific software operating environments, application domains, and other characteristics of interest;
- Improving the DEGM8SV by searching for and collecting additional data (including SDLCBegin / SwAccept pairs) and rerunning the regressions;
- Creating specific growth model instances regressed from the same data subset that is regressed to create specific Joint Cost and Duration Estimating Relationships, thus creating unified software estimating methodologies that are specific to certain types of software.
**Appendix A**  Custom CDFs for DEGM8 Default and DEGM8SV

### DEGM8 Default

<table>
<thead>
<tr>
<th>Percentile</th>
<th>Value</th>
<th>Percentile</th>
<th>Value</th>
<th>Percentile</th>
<th>Value</th>
<th>Percentile</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>JCDER349 e_GN_CDF</td>
<td>0.21645022</td>
<td>JCDER349 e_GM_CDF</td>
<td>0.35211268</td>
<td>JCDER349 e_GU_CDF</td>
<td>0.33783784</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.09821762</td>
<td>0.10673836</td>
<td>0.09492565</td>
<td>0.10941358</td>
<td>0.09673836</td>
<td>0.10673836</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.64935065</td>
<td>0.10232146</td>
<td>1.05633803</td>
<td>0.10684445</td>
<td>1.01351351</td>
<td>0.10684445</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.08225108</td>
<td>0.11109953</td>
<td>1.76056338</td>
<td>0.12369618</td>
<td>1.68918919</td>
<td>0.12369618</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.51515152</td>
<td>0.13489884</td>
<td>2.81690141</td>
<td>0.15941357</td>
<td>2.36486846</td>
<td>0.15941357</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.94805195</td>
<td>0.16370668</td>
<td>3.78323944</td>
<td>0.17850355</td>
<td>3.71621622</td>
<td>0.17850355</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.38095238</td>
<td>0.18335752</td>
<td>4.76056338</td>
<td>0.19543874</td>
<td>4.39189189</td>
<td>0.19543874</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.81385281</td>
<td>0.19303899</td>
<td>5.73746479</td>
<td>0.20389943</td>
<td>5.06756757</td>
<td>0.20389943</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.24675325</td>
<td>0.21836560</td>
<td>6.71621622</td>
<td>0.21836560</td>
<td>5.98591549</td>
<td>0.21836560</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.67965368</td>
<td>0.23557328</td>
<td>7.68918919</td>
<td>0.23714712</td>
<td>7.09459459</td>
<td>0.23714712</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.11254411</td>
<td>0.25017776</td>
<td>8.68918919</td>
<td>0.25017776</td>
<td>6.69014085</td>
<td>0.25017776</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.54544554</td>
<td>0.26583739</td>
<td>9.68918919</td>
<td>0.26583739</td>
<td>6.27705628</td>
<td>0.26583739</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.97835498</td>
<td>0.27952010</td>
<td>10.68918919</td>
<td>0.27952010</td>
<td>6.27705628</td>
<td>0.27952010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.41125541</td>
<td>0.29379456</td>
<td>11.68918919</td>
<td>0.29379456</td>
<td>6.27705628</td>
<td>0.29379456</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.84415584</td>
<td>0.30705405</td>
<td>12.68918919</td>
<td>0.30705405</td>
<td>6.27705628</td>
<td>0.30705405</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.27705628</td>
<td>0.32026359</td>
<td>13.68918919</td>
<td>0.32026359</td>
<td>6.27705628</td>
<td>0.32026359</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.70995671</td>
<td>0.33347309</td>
<td>14.68918919</td>
<td>0.33347309</td>
<td>6.27705628</td>
<td>0.33347309</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.14285714</td>
<td>0.34668253</td>
<td>15.68918919</td>
<td>0.34668253</td>
<td>6.27705628</td>
<td>0.34668253</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.57575758</td>
<td>0.36009178</td>
<td>16.68918919</td>
<td>0.36009178</td>
<td>6.27705628</td>
<td>0.36009178</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.00865801</td>
<td>0.37350102</td>
<td>17.68918919</td>
<td>0.37350102</td>
<td>6.27705628</td>
<td>0.37350102</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.44155844</td>
<td>0.38701026</td>
<td>18.68918919</td>
<td>0.38701026</td>
<td>6.27705628</td>
<td>0.38701026</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.87445887</td>
<td>0.40051950</td>
<td>19.68918919</td>
<td>0.40051950</td>
<td>6.27705628</td>
<td>0.40051950</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.30735931</td>
<td>0.41401874</td>
<td>20.68918919</td>
<td>0.41401874</td>
<td>6.27705628</td>
<td>0.41401874</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.74025974</td>
<td>0.42751798</td>
<td>21.68918919</td>
<td>0.42751798</td>
<td>6.27705628</td>
<td>0.42751798</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.38961039</td>
<td>0.44101722</td>
<td>22.68918919</td>
<td>0.44101722</td>
<td>6.27705628</td>
<td>0.44101722</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.83901404</td>
<td>0.45451646</td>
<td>23.68918919</td>
<td>0.45451646</td>
<td>6.27705628</td>
<td>0.45451646</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.38841769</td>
<td>0.46801570</td>
<td>24.68918919</td>
<td>0.46801570</td>
<td>6.27705628</td>
<td>0.46801570</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.93782134</td>
<td>0.48151494</td>
<td>25.68918919</td>
<td>0.48151494</td>
<td>6.27705628</td>
<td>0.48151494</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.48722500</td>
<td>0.49501418</td>
<td>26.68918919</td>
<td>0.49501418</td>
<td>6.27705628</td>
<td>0.49501418</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.03662865</td>
<td>0.50851342</td>
<td>27.68918919</td>
<td>0.50851342</td>
<td>6.27705628</td>
<td>0.50851342</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentile Value</td>
<td>Percentile Value</td>
<td>Percentile Value</td>
<td>Percentile Value</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>------------------</td>
<td>------------------</td>
<td>------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17.53246753</td>
<td>0.49222882</td>
<td>28.52112676</td>
<td>0.73552024</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17.96536797</td>
<td>0.49240754</td>
<td>29.22535211</td>
<td>0.77890545</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18.39826840</td>
<td>0.50098235</td>
<td>29.92957746</td>
<td>0.78491907</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18.83116883</td>
<td>0.50473046</td>
<td>30.63308202</td>
<td>0.79861919</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19.26406926</td>
<td>0.50505288</td>
<td>31.33808217</td>
<td>0.81071575</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19.69696970</td>
<td>0.51629065</td>
<td>32.04225352</td>
<td>0.84507406</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20.34632036</td>
<td>0.53237301</td>
<td>32.74647887</td>
<td>0.86474552</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20.95671000</td>
<td>0.53939074</td>
<td>33.45070423</td>
<td>0.88659754</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21.42857143</td>
<td>0.53893907</td>
<td>34.15492958</td>
<td>0.90125044</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21.86147186</td>
<td>0.54367058</td>
<td>34.85915493</td>
<td>0.90723316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22.29437229</td>
<td>0.54517965</td>
<td>35.56338028</td>
<td>0.90723316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22.72727273</td>
<td>0.55088599</td>
<td>36.26760563</td>
<td>0.90723316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23.16017316</td>
<td>0.56633499</td>
<td>36.97183099</td>
<td>0.90723316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23.59307595</td>
<td>0.58451797</td>
<td>37.67605639</td>
<td>0.90723316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24.02597403</td>
<td>0.59949251</td>
<td>38.38083099</td>
<td>0.90723316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24.45887446</td>
<td>0.60637477</td>
<td>39.08450704</td>
<td>0.90723316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24.89177489</td>
<td>0.63906530</td>
<td>39.78873239</td>
<td>0.90723316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25.32467532</td>
<td>0.64736349</td>
<td>40.49295775</td>
<td>0.90723316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25.75757576</td>
<td>0.64955919</td>
<td>41.19718310</td>
<td>0.90723316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26.19047619</td>
<td>0.65069836</td>
<td>41.89140845</td>
<td>0.90723316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26.62337662</td>
<td>0.64983665</td>
<td>42.59467832</td>
<td>0.90723316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27.05627706</td>
<td>0.66313574</td>
<td>43.29890148</td>
<td>0.90723316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27.48917749</td>
<td>0.68437058</td>
<td>44.00140845</td>
<td>0.90723316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27.92207792</td>
<td>0.70634203</td>
<td>44.71309306</td>
<td>0.90723316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28.35497835</td>
<td>0.71032622</td>
<td>45.42535211</td>
<td>0.90723316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28.78787879</td>
<td>0.71211396</td>
<td>46.13760563</td>
<td>0.90723316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29.22079222</td>
<td>0.72495221</td>
<td>46.83989592</td>
<td>0.90723316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29.6567965</td>
<td>0.76954933</td>
<td>47.53321271</td>
<td>0.90723316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30.08658009</td>
<td>0.78069905</td>
<td>48.23943662</td>
<td>0.90723316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30.51948052</td>
<td>0.82860624</td>
<td>49.04788732</td>
<td>0.90723316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30.95238095</td>
<td>0.84004617</td>
<td>50.35211268</td>
<td>0.90723316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31.36017316</td>
<td>0.83994949</td>
<td>51.05633803</td>
<td>0.90723316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31.60173161</td>
<td>0.83994949</td>
<td>51.76056333</td>
<td>0.90723316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32.25108225</td>
<td>0.84355358</td>
<td>52.46478873</td>
<td>0.90723316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32.68939828</td>
<td>0.84804201</td>
<td>53.16901408</td>
<td>0.90723316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33.11683122</td>
<td>0.84863834</td>
<td>53.87323944</td>
<td>0.90723316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33.54978355</td>
<td>0.85737260</td>
<td>54.57746479</td>
<td>0.90723316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33.98263938</td>
<td>0.86288518</td>
<td>55.28169014</td>
<td>0.90723316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34.41558442</td>
<td>0.86565232</td>
<td>56.04225353</td>
<td>0.90723316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34.84848485</td>
<td>0.88588056</td>
<td>57.04225353</td>
<td>0.90723316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentile Value</td>
<td>Percentile Value</td>
<td>Percentile Value</td>
<td>Percentile Value</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>------------------</td>
<td>------------------</td>
<td>------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35.28138528</td>
<td>0.8962728</td>
<td>57.04225354</td>
<td>1.1944252</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35.71428571</td>
<td>0.8902808</td>
<td>57.04225355</td>
<td>1.1944253</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36.14718615</td>
<td>0.8959510</td>
<td>58.80281690</td>
<td>1.21153873</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36.58008658</td>
<td>0.9028560</td>
<td>59.50704225</td>
<td>1.22300063</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37.01287012</td>
<td>0.90634077</td>
<td>60.21126761</td>
<td>1.23646399</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37.44588745</td>
<td>0.9154061</td>
<td>60.91549296</td>
<td>1.25008616</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37.87878788</td>
<td>0.92323309</td>
<td>61.61971831</td>
<td>1.25483924</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38.31188331</td>
<td>0.93631473</td>
<td>62.32394366</td>
<td>1.29524374</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38.74458874</td>
<td>0.94482377</td>
<td>63.02816901</td>
<td>1.31605101</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39.17748918</td>
<td>0.95128544</td>
<td>63.73239437</td>
<td>1.32506191</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39.61038961</td>
<td>0.96145076</td>
<td>64.43661972</td>
<td>1.34117075</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40.04329004</td>
<td>0.98448707</td>
<td>65.49295775</td>
<td>1.35439267</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40.47619048</td>
<td>1.01076477</td>
<td>65.94929576</td>
<td>1.35439268</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40.90909091</td>
<td>1.02921046</td>
<td>66.54929577</td>
<td>1.37658361</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41.34199134</td>
<td>1.03783072</td>
<td>67.25352113</td>
<td>1.37915589</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41.77489177</td>
<td>1.04729221</td>
<td>67.95774648</td>
<td>1.40701851</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42.20779221</td>
<td>1.05118515</td>
<td>68.66197183</td>
<td>1.41422111</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42.64069264</td>
<td>1.05132735</td>
<td>69.36619718</td>
<td>1.42637124</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43.07359307</td>
<td>1.06718546</td>
<td>70.07042254</td>
<td>1.44531403</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43.50649351</td>
<td>1.07451964</td>
<td>70.77464789</td>
<td>1.47315839</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43.93939394</td>
<td>1.07659579</td>
<td>71.47887324</td>
<td>1.49081150</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44.37229437</td>
<td>1.08163376</td>
<td>72.18309869</td>
<td>1.54078991</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44.80519481</td>
<td>1.08471770</td>
<td>72.88732394</td>
<td>1.55539476</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45.23809524</td>
<td>1.08922249</td>
<td>73.59154930</td>
<td>1.56382122</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45.67099567</td>
<td>1.09514976</td>
<td>74.29577465</td>
<td>1.56491152</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46.10389610</td>
<td>1.09565189</td>
<td>75.00000000</td>
<td>1.56797051</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46.53679654</td>
<td>1.09800178</td>
<td>75.70422535</td>
<td>1.58043573</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46.96999697</td>
<td>1.10384411</td>
<td>76.40845070</td>
<td>1.62531725</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47.40259740</td>
<td>1.11353633</td>
<td>77.11267606</td>
<td>1.78452498</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47.83549784</td>
<td>1.12682812</td>
<td>77.81690141</td>
<td>1.80415794</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48.26839827</td>
<td>1.12748891</td>
<td>78.52112676</td>
<td>1.83920114</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48.70129870</td>
<td>1.13572557</td>
<td>79.22532511</td>
<td>1.88882286</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>49.13419913</td>
<td>1.13830528</td>
<td>79.92957746</td>
<td>1.96622902</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>49.56709957</td>
<td>1.15472812</td>
<td>80.63380282</td>
<td>1.97495992</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50.00000000</td>
<td>1.16895017</td>
<td>81.33808217</td>
<td>2.17124647</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50.43290043</td>
<td>1.17902924</td>
<td>82.04225352</td>
<td>2.23511451</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50.86580087</td>
<td>1.18763087</td>
<td>82.74648787</td>
<td>2.31668772</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>51.29870130</td>
<td>1.19559110</td>
<td>83.45070423</td>
<td>2.33664047</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>51.73160173</td>
<td>1.20846700</td>
<td>84.15492958</td>
<td>2.42965630</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>52.16450216</td>
<td>1.21009716</td>
<td>84.85915493</td>
<td>2.52126116</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentile</td>
<td>Value</td>
<td>Percentile</td>
<td>Value</td>
<td>Percentile</td>
<td>Value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>-------</td>
<td>------------</td>
<td>-------</td>
<td>------------</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JCDER349_e_GN_CDF</td>
<td>JCDER349_e_GM_CDF</td>
<td>JCDER349_e_GU_CDF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>52.59740260</td>
<td>1.21718712</td>
<td>85.56338028</td>
<td>3.94768260</td>
<td>82.09459459</td>
<td>1.42239106</td>
<td></td>
<td></td>
</tr>
<tr>
<td>53.67965368</td>
<td>1.21856641</td>
<td>86.26760563</td>
<td>4.03681511</td>
<td>82.77027027</td>
<td>1.47491346</td>
<td></td>
<td></td>
</tr>
<tr>
<td>53.67965369</td>
<td>1.21856642</td>
<td>86.97183099</td>
<td>4.08535697</td>
<td>83.44594595</td>
<td>1.54087554</td>
<td></td>
<td></td>
</tr>
<tr>
<td>53.67965370</td>
<td>1.21856643</td>
<td>87.67605634</td>
<td>4.31916018</td>
<td>84.12162162</td>
<td>1.56165460</td>
<td></td>
<td></td>
</tr>
<tr>
<td>53.67965371</td>
<td>1.21856644</td>
<td>88.38028169</td>
<td>4.38631400</td>
<td>84.79729730</td>
<td>1.63760872</td>
<td></td>
<td></td>
</tr>
<tr>
<td>54.76190476</td>
<td>1.22070022</td>
<td>89.08450704</td>
<td>4.46549480</td>
<td>85.47297297</td>
<td>1.64581226</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55.19480519</td>
<td>1.24888022</td>
<td>89.78873239</td>
<td>4.69566894</td>
<td>86.14864865</td>
<td>1.74727755</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55.62770563</td>
<td>1.26316574</td>
<td>90.49295775</td>
<td>5.36466737</td>
<td>86.82432432</td>
<td>1.85308025</td>
<td></td>
<td></td>
</tr>
<tr>
<td>56.06060606</td>
<td>1.30718866</td>
<td>91.19718310</td>
<td>5.64451232</td>
<td>87.50000000</td>
<td>1.96463749</td>
<td></td>
<td></td>
</tr>
<tr>
<td>56.49350649</td>
<td>1.30311379</td>
<td>91.90140845</td>
<td>6.15092302</td>
<td>88.17567568</td>
<td>2.19879720</td>
<td></td>
<td></td>
</tr>
<tr>
<td>56.92640693</td>
<td>1.30328666</td>
<td>92.60563380</td>
<td>6.60824135</td>
<td>88.85135135</td>
<td>2.25059828</td>
<td></td>
<td></td>
</tr>
<tr>
<td>57.35930736</td>
<td>1.30469479</td>
<td>93.30985915</td>
<td>6.89342132</td>
<td>89.52702703</td>
<td>2.30862992</td>
<td></td>
<td></td>
</tr>
<tr>
<td>57.79220779</td>
<td>1.31474442</td>
<td>94.01408451</td>
<td>6.93822543</td>
<td>90.20270270</td>
<td>2.45242958</td>
<td></td>
<td></td>
</tr>
<tr>
<td>58.22510823</td>
<td>1.32808421</td>
<td>94.71830986</td>
<td>8.08527304</td>
<td>90.87837838</td>
<td>2.49114222</td>
<td></td>
<td></td>
</tr>
<tr>
<td>58.65800866</td>
<td>1.33893139</td>
<td>95.42253521</td>
<td>8.68058303</td>
<td>91.55404505</td>
<td>2.56580767</td>
<td></td>
<td></td>
</tr>
<tr>
<td>59.09090909</td>
<td>1.34017032</td>
<td>96.12686562</td>
<td>9.22972973</td>
<td>92.22972973</td>
<td>2.84820035</td>
<td></td>
<td></td>
</tr>
<tr>
<td>59.52380952</td>
<td>1.36144435</td>
<td>96.85940435</td>
<td>9.20540541</td>
<td>92.90540541</td>
<td>3.42449546</td>
<td></td>
<td></td>
</tr>
<tr>
<td>59.95670996</td>
<td>1.36740683</td>
<td>97.59173209</td>
<td>93.58108108</td>
<td>93.58108108</td>
<td>4.36850399</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60.38961039</td>
<td>1.40539438</td>
<td>98.32513402</td>
<td>94.25675676</td>
<td>94.25675676</td>
<td>4.58082833</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60.82251082</td>
<td>1.40738383</td>
<td>99.05853402</td>
<td>94.93242343</td>
<td>94.93242343</td>
<td>6.70576472</td>
<td></td>
<td></td>
</tr>
<tr>
<td>61.25541126</td>
<td>1.40801947</td>
<td>99.79193402</td>
<td>95.60810811</td>
<td>95.60810811</td>
<td>6.98289033</td>
<td></td>
<td></td>
</tr>
<tr>
<td>61.68831169</td>
<td>1.40953550</td>
<td>1.00541126</td>
<td>96.3229135</td>
<td>1.00541126</td>
<td>6.98289033</td>
<td></td>
<td></td>
</tr>
<tr>
<td>62.12121212</td>
<td>1.41229135</td>
<td>1.01251126</td>
<td>1.02075900</td>
<td>1.02075900</td>
<td>6.98289033</td>
<td></td>
<td></td>
</tr>
<tr>
<td>62.55411255</td>
<td>1.42075900</td>
<td>1.02801947</td>
<td>1.43790840</td>
<td>1.43790840</td>
<td>4.4779958</td>
<td></td>
<td></td>
</tr>
<tr>
<td>63.08715392</td>
<td>1.44077958</td>
<td>1.44130955</td>
<td>6.38521385</td>
<td>6.38521385</td>
<td>4.4779958</td>
<td></td>
<td></td>
</tr>
<tr>
<td>63.41991342</td>
<td>1.44795403</td>
<td>6.426571429</td>
<td>1.44795403</td>
<td>1.44795403</td>
<td>6.426571429</td>
<td></td>
<td></td>
</tr>
<tr>
<td>63.58441558</td>
<td>1.45670678</td>
<td>6.47891775</td>
<td>1.47756225</td>
<td>6.47891775</td>
<td>1.47756225</td>
<td></td>
<td></td>
</tr>
<tr>
<td>64.10131602</td>
<td>1.46447619</td>
<td>6.60167132</td>
<td>1.47967037</td>
<td>6.60167132</td>
<td>1.47967037</td>
<td></td>
<td></td>
</tr>
<tr>
<td>64.45021645</td>
<td>1.46447619</td>
<td>6.68316688</td>
<td>1.47076292</td>
<td>6.68316688</td>
<td>1.47076292</td>
<td></td>
<td></td>
</tr>
<tr>
<td>64.80021645</td>
<td>1.48139681</td>
<td>6.73160173</td>
<td>1.47967037</td>
<td>6.73160173</td>
<td>1.47967037</td>
<td></td>
<td></td>
</tr>
<tr>
<td>64.81471861</td>
<td>1.48462273</td>
<td>6.774891775</td>
<td>1.47756225</td>
<td>6.774891775</td>
<td>1.47756225</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65.04761905</td>
<td>1.48575596</td>
<td>6.818181818</td>
<td>1.48139681</td>
<td>6.818181818</td>
<td>1.48139681</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65.48051948</td>
<td>1.49090125</td>
<td>6.861471861</td>
<td>1.48462273</td>
<td>6.861471861</td>
<td>1.48462273</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65.92380952</td>
<td>1.49829033</td>
<td>6.904761905</td>
<td>1.48575596</td>
<td>6.904761905</td>
<td>1.48575596</td>
<td></td>
<td></td>
</tr>
<tr>
<td>66.36580066</td>
<td>1.50629033</td>
<td>6.948051948</td>
<td>1.49090125</td>
<td>6.948051948</td>
<td>1.49090125</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentile</td>
<td>Value</td>
<td>Percentile</td>
<td>Value</td>
<td>Percentile</td>
<td>Value</td>
<td>Percentile</td>
<td>Value</td>
</tr>
<tr>
<td>------------</td>
<td>--------</td>
<td>------------</td>
<td>--------</td>
<td>------------</td>
<td>--------</td>
<td>------------</td>
<td>--------</td>
</tr>
<tr>
<td>JCDER349_e_GN_CDF</td>
<td>69.91341991</td>
<td>1.49141098</td>
<td>JCDER349_e_GM_CDF</td>
<td>70.34632035</td>
<td>1.49472454</td>
<td>JCDER349_e_GU_CDF</td>
<td>70.77922078</td>
</tr>
<tr>
<td>JCDER349_e_GN_CDF</td>
<td>71.21212121</td>
<td>1.50360791</td>
<td>JCDER349_e_GM_CDF</td>
<td>71.64502165</td>
<td>1.50953145</td>
<td>JCDER349_e_GU_CDF</td>
<td>72.07792208</td>
</tr>
<tr>
<td>JCDER349_e_GN_CDF</td>
<td>72.51082251</td>
<td>1.51238818</td>
<td>JCDER349_e_GM_CDF</td>
<td>72.94372294</td>
<td>1.51418200</td>
<td>JCDER349_e_GU_CDF</td>
<td>73.37662338</td>
</tr>
<tr>
<td>JCDER349_e_GN_CDF</td>
<td>73.80952381</td>
<td>1.52604731</td>
<td>JCDER349_e_GM_CDF</td>
<td>74.24242424</td>
<td>1.52724671</td>
<td>JCDER349_e_GU_CDF</td>
<td>74.67532468</td>
</tr>
<tr>
<td>JCDER349_e_GN_CDF</td>
<td>75.10822511</td>
<td>1.54909324</td>
<td>JCDER349_e_GM_CDF</td>
<td>75.54112554</td>
<td>1.55196396</td>
<td>JCDER349_e_GU_CDF</td>
<td>75.97402597</td>
</tr>
<tr>
<td>JCDER349_e_GN_CDF</td>
<td>76.40692641</td>
<td>1.60155165</td>
<td>JCDER349_e_GM_CDF</td>
<td>76.83982684</td>
<td>1.61218716</td>
<td>JCDER349_e_GU_CDF</td>
<td>77.27272727</td>
</tr>
<tr>
<td>JCDER349_e_GN_CDF</td>
<td>77.70562771</td>
<td>1.63103661</td>
<td>JCDER349_e_GM_CDF</td>
<td>78.13852814</td>
<td>1.63662640</td>
<td>JCDER349_e_GU_CDF</td>
<td>78.57142857</td>
</tr>
<tr>
<td>JCDER349_e_GN_CDF</td>
<td>79.00432900</td>
<td>1.63919405</td>
<td>JCDER349_e_GM_CDF</td>
<td>79.43722944</td>
<td>1.65367135</td>
<td>JCDER349_e_GU_CDF</td>
<td>79.87012987</td>
</tr>
<tr>
<td>JCDER349_e_GN_CDF</td>
<td>80.30303030</td>
<td>1.71500546</td>
<td>JCDER349_e_GM_CDF</td>
<td>80.73593074</td>
<td>1.72094563</td>
<td>JCDER349_e_GU_CDF</td>
<td>81.16883117</td>
</tr>
<tr>
<td>JCDER349_e_GN_CDF</td>
<td>81.60173160</td>
<td>1.75624422</td>
<td>JCDER349_e_GM_CDF</td>
<td>82.03463203</td>
<td>1.79557587</td>
<td>JCDER349_e_GU_CDF</td>
<td>82.46753247</td>
</tr>
<tr>
<td>JCDER349_e_GN_CDF</td>
<td>82.90043290</td>
<td>1.86969853</td>
<td>JCDER349_e_GM_CDF</td>
<td>83.33333333</td>
<td>2.00941077</td>
<td>JCDER349_e_GU_CDF</td>
<td>83.76623377</td>
</tr>
<tr>
<td>JCDER349_e_GN_CDF</td>
<td>84.19913420</td>
<td>2.03365666</td>
<td>JCDER349_e_GM_CDF</td>
<td>84.63203463</td>
<td>2.10800121</td>
<td>JCDER349_e_GU_CDF</td>
<td>85.06498506</td>
</tr>
<tr>
<td>JCDER349_e_GN_CDF</td>
<td>85.49783550</td>
<td>2.26979724</td>
<td>JCDER349_e_GM_CDF</td>
<td>85.93073593</td>
<td>2.42286591</td>
<td>JCDER349_e_GU_CDF</td>
<td>86.36363636</td>
</tr>
<tr>
<td>JCDER349_e_GN_CDF</td>
<td>86.79653680</td>
<td>2.43417811</td>
<td>JCDER349_e_GM_CDF</td>
<td>87.22942942</td>
<td>2.48624510</td>
<td>JCDER349_e_GU_CDF</td>
<td>87.66233334</td>
</tr>
<tr>
<td>JCDER349_e_GN_CDF</td>
<td>88.09523336</td>
<td>2.67284370</td>
<td>JCDER349_e_GM_CDF</td>
<td>88.52813636</td>
<td>2.76442443</td>
<td>JCDER349_e_GU_CDF</td>
<td>88.96013939</td>
</tr>
<tr>
<td>JCDER349_e_GN_CDF</td>
<td>89.39293243</td>
<td>2.91982545</td>
<td>JCDER349_e_GM_CDF</td>
<td>89.82583543</td>
<td>3.07804726</td>
<td>JCDER349_e_GU_CDF</td>
<td>90.25873846</td>
</tr>
<tr>
<td>JCDER349_e_GN_CDF</td>
<td>90.69163153</td>
<td>3.34552107</td>
<td>JCDER349_e_GM_CDF</td>
<td>91.12453453</td>
<td>3.54754288</td>
<td>JCDER349_e_GU_CDF</td>
<td>91.55743756</td>
</tr>
<tr>
<td>JCDER349_e_GN_CDF</td>
<td>91.95033463</td>
<td>3.98908647</td>
<td>JCDER349_e_GM_CDF</td>
<td>92.38323763</td>
<td>4.32410828</td>
<td>JCDER349_e_GU_CDF</td>
<td>92.81614066</td>
</tr>
<tr>
<td>JCDER349_e_GN_CDF</td>
<td>93.18893773</td>
<td>5.24213420</td>
<td>JCDER349_e_GM_CDF</td>
<td>93.62184073</td>
<td>5.75415601</td>
<td>JCDER349_e_GU_CDF</td>
<td>94.05474374</td>
</tr>
<tr>
<td>JCDER349_e_GN_CDF</td>
<td>94.36363636</td>
<td>7.70290594</td>
<td>JCDER349_e_GM_CDF</td>
<td>94.89653936</td>
<td>9.65292775</td>
<td>JCDER349_e_GU_CDF</td>
<td>95.32944237</td>
</tr>
<tr>
<td>JCDER349_e_GN_CDF</td>
<td>95.51082251</td>
<td>12.64175138</td>
<td>JCDER349_e_GM_CDF</td>
<td>96.04362552</td>
<td>14.79167320</td>
<td>JCDER349_e_GU_CDF</td>
<td>96.47642854</td>
</tr>
<tr>
<td>JCDER349_e_GN_CDF</td>
<td>96.71212121</td>
<td>17.04159510</td>
<td>JCDER349_e_GM_CDF</td>
<td>97.24492424</td>
<td>19.49151692</td>
<td>JCDER349_e_GU_CDF</td>
<td>97.67772747</td>
</tr>
<tr>
<td>JCDER349_e_GN_CDF</td>
<td>97.87922087</td>
<td>22.14143873</td>
<td>JCDER349_e_GM_CDF</td>
<td>98.41102399</td>
<td>24.99136055</td>
<td>JCDER349_e_GU_CDF</td>
<td>98.84382722</td>
</tr>
<tr>
<td>JCDER349_e_GN_CDF</td>
<td>99.04362552</td>
<td>28.19128236</td>
<td>JCDER349_e_GM_CDF</td>
<td>99.57542866</td>
<td>32.74120418</td>
<td>JCDER349_e_GU_CDF</td>
<td>100.00000000</td>
</tr>
</tbody>
</table>

Presented at the 2018 ICEAA Professional Development & Training Workshop - www.iceaaonline.com
<table>
<thead>
<tr>
<th>Percentile</th>
<th>Value</th>
<th>Percentile</th>
<th>Value</th>
<th>Percentile</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>JCDER349_e_GN_CDF</td>
<td></td>
<td>JCDER349_e_GM_CDF</td>
<td></td>
<td>JCDER349_e_GU_CDF</td>
<td></td>
</tr>
<tr>
<td>87.22943723</td>
<td>2.53313064</td>
<td>87.66233766</td>
<td>2.68118931</td>
<td>88.09523810</td>
<td>2.81748607</td>
</tr>
<tr>
<td>88.52813853</td>
<td>2.83066175</td>
<td>88.96103896</td>
<td>2.88697964</td>
<td>89.39939393</td>
<td>3.02215573</td>
</tr>
<tr>
<td>89.82683983</td>
<td>3.07194389</td>
<td>90.25974026</td>
<td>3.14832918</td>
<td>90.69264069</td>
<td>3.17650058</td>
</tr>
<tr>
<td>91.12554113</td>
<td>3.33077404</td>
<td>91.55844156</td>
<td>3.61764732</td>
<td>91.99134199</td>
<td>3.82766747</td>
</tr>
<tr>
<td>92.42424242</td>
<td>3.85337172</td>
<td>92.85714286</td>
<td>3.87914435</td>
<td>93.29004329</td>
<td>4.20076391</td>
</tr>
<tr>
<td>93.72294372</td>
<td>4.31303466</td>
<td>94.15584416</td>
<td>4.53838753</td>
<td>94.58744594</td>
<td>4.99452737</td>
</tr>
<tr>
<td>95.02164502</td>
<td>5.16629480</td>
<td>95.45454545</td>
<td>5.48454971</td>
<td>95.88744589</td>
<td>5.68416997</td>
</tr>
<tr>
<td>96.32034632</td>
<td>6.35782957</td>
<td>96.75324675</td>
<td>7.35600544</td>
<td>97.18614719</td>
<td>8.80095370</td>
</tr>
</tbody>
</table>
## DEGM8SV

<table>
<thead>
<tr>
<th>Percentile</th>
<th>Value</th>
<th>Percentile</th>
<th>Value</th>
<th>Percentile</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>JCDER101 e_GN_CDF</strong></td>
<td><strong>JCDER101 e_GM_CDF</strong></td>
<td><strong>JCDER101 e_GU_CDF</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.42857143</td>
<td>0.39797761</td>
<td>2.17391304</td>
<td>0.12594765</td>
<td>1.61290323</td>
<td>0.13570110</td>
</tr>
<tr>
<td>4.28571429</td>
<td>0.44806233</td>
<td>6.52173913</td>
<td>0.15264906</td>
<td>4.83709688</td>
<td>0.13926041</td>
</tr>
<tr>
<td>7.14285714</td>
<td>0.45520488</td>
<td>10.86956522</td>
<td>0.29956749</td>
<td>8.06451613</td>
<td>0.15562935</td>
</tr>
<tr>
<td>10.00000000</td>
<td>0.49531455</td>
<td>15.21739130</td>
<td>0.30107109</td>
<td>11.29032258</td>
<td>0.15998283</td>
</tr>
<tr>
<td>12.85714286</td>
<td>0.53769705</td>
<td>19.56521739</td>
<td>0.51047319</td>
<td>14.51612903</td>
<td>0.17382185</td>
</tr>
<tr>
<td>15.71428571</td>
<td>0.59597535</td>
<td>23.91304348</td>
<td>0.52432572</td>
<td>17.74193548</td>
<td>0.45946000</td>
</tr>
<tr>
<td>18.57142857</td>
<td>0.63250633</td>
<td>28.26086957</td>
<td>0.56496728</td>
<td>20.96774194</td>
<td>0.46209723</td>
</tr>
<tr>
<td>21.42857143</td>
<td>0.64519672</td>
<td>32.60869565</td>
<td>0.56835342</td>
<td>24.19354839</td>
<td>0.46922325</td>
</tr>
<tr>
<td>24.28571429</td>
<td>0.70538335</td>
<td>36.95652174</td>
<td>0.58355543</td>
<td>27.41935484</td>
<td>0.50365112</td>
</tr>
<tr>
<td>27.14285714</td>
<td>0.85185971</td>
<td>41.30434783</td>
<td>0.61270300</td>
<td>30.64516129</td>
<td>0.51921489</td>
</tr>
<tr>
<td>30.00000000</td>
<td>0.90596191</td>
<td>45.65217391</td>
<td>1.10082647</td>
<td>33.87096774</td>
<td>0.61099536</td>
</tr>
<tr>
<td>32.85714286</td>
<td>0.96374280</td>
<td>50.00000000</td>
<td>1.12965659</td>
<td>37.09677419</td>
<td>0.65629299</td>
</tr>
<tr>
<td>35.71428571</td>
<td>1.02491388</td>
<td>54.34782609</td>
<td>1.26664643</td>
<td>40.32280656</td>
<td>0.68347624</td>
</tr>
<tr>
<td>38.57142857</td>
<td>1.04560007</td>
<td>60.86956522</td>
<td>1.49520849</td>
<td>43.54838710</td>
<td>0.70730690</td>
</tr>
<tr>
<td>41.42857143</td>
<td>1.07067023</td>
<td>66.89565236</td>
<td>1.49520850</td>
<td>46.77419355</td>
<td>0.72038296</td>
</tr>
<tr>
<td>44.28571429</td>
<td>1.07451465</td>
<td>67.39130435</td>
<td>1.62277675</td>
<td>50.00000000</td>
<td>0.77768603</td>
</tr>
<tr>
<td>47.14285714</td>
<td>1.11223696</td>
<td>71.73913043</td>
<td>1.64009976</td>
<td>53.22580645</td>
<td>0.79931977</td>
</tr>
<tr>
<td>50.00000000</td>
<td>1.12029311</td>
<td>78.26086957</td>
<td>1.74582811</td>
<td>56.45161290</td>
<td>1.25171660</td>
</tr>
<tr>
<td>52.85714286</td>
<td>1.12912742</td>
<td>78.26086958</td>
<td>1.74582812</td>
<td>59.67741935</td>
<td>1.30113261</td>
</tr>
<tr>
<td>55.71428571</td>
<td>1.14227580</td>
<td>84.78260870</td>
<td>1.85588735</td>
<td>62.90322581</td>
<td>1.33957377</td>
</tr>
<tr>
<td>58.57142857</td>
<td>1.15615324</td>
<td>89.13043478</td>
<td>6.12074288</td>
<td>66.12903226</td>
<td>1.47613209</td>
</tr>
<tr>
<td>61.42857143</td>
<td>1.15739641</td>
<td>93.47826087</td>
<td>7.37744324</td>
<td>69.35438781</td>
<td>1.53213683</td>
</tr>
<tr>
<td>64.28571429</td>
<td>1.16677484</td>
<td>97.82606069</td>
<td>7.83946109</td>
<td>72.58064516</td>
<td>2.05349824</td>
</tr>
<tr>
<td>67.14285714</td>
<td>1.20762787</td>
<td>75.80645161</td>
<td>2.13644652</td>
<td>79.03225806</td>
<td>2.35487073</td>
</tr>
<tr>
<td>70.00000000</td>
<td>1.22701029</td>
<td>82.25806452</td>
<td>3.77161536</td>
<td>75.14285714</td>
<td>2.51878929</td>
</tr>
<tr>
<td>72.85714286</td>
<td>1.23870624</td>
<td>85.48370979</td>
<td>4.52797607</td>
<td>78.57142857</td>
<td>1.29971870</td>
</tr>
<tr>
<td>75.14285714</td>
<td>1.25878929</td>
<td>88.70967742</td>
<td>5.73543246</td>
<td>81.42857143</td>
<td>1.35686365</td>
</tr>
<tr>
<td>84.28571429</td>
<td>1.37351749</td>
<td>95.16129032</td>
<td>10.07136883</td>
<td>87.14285714</td>
<td>1.41537438</td>
</tr>
<tr>
<td>90.00000000</td>
<td>1.51039099</td>
<td>98.38709677</td>
<td>12.94622288</td>
<td>92.85714286</td>
<td>1.53898881</td>
</tr>
<tr>
<td>95.71428571</td>
<td>1.63834787</td>
<td>98.57142857</td>
<td>3.56984543</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Appendix B  DEGM8 Regressions Summary

<table>
<thead>
<tr>
<th>Source Database:</th>
<th>SRDR 2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filter Criteria:</td>
<td>SerialNo2015: &gt;0; Report: 2630-3; Si: TRUE; Nonphysical: TRUE; GF3V: TRUE</td>
</tr>
<tr>
<td>ESLOC Calculation Method:</td>
<td>Popp 2015</td>
</tr>
<tr>
<td>Custom Select:</td>
<td>All</td>
</tr>
<tr>
<td>DSLOC Measurement Point:</td>
<td>Final</td>
</tr>
<tr>
<td>Effort SDLC Scope:</td>
<td>Core</td>
</tr>
<tr>
<td>Duration SDLC Scope:</td>
<td>Core</td>
</tr>
</tbody>
</table>

### Data Set Statistics

- **Number of Data Points (observations):** 578

<table>
<thead>
<tr>
<th></th>
<th>Smallest</th>
<th>Largest</th>
<th>Mean</th>
<th>Geomean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size (ESLOC):</td>
<td>12</td>
<td>913,602</td>
<td>70,172</td>
<td>24,859</td>
</tr>
<tr>
<td>Core Effort (pm):</td>
<td>0.1</td>
<td>9,526.9</td>
<td>421.1</td>
<td>129.8</td>
</tr>
<tr>
<td>Core Duration (cm):</td>
<td>1.15</td>
<td>115.02</td>
<td>35.59</td>
<td>28.71</td>
</tr>
<tr>
<td>All-All Productivity (ESLOC / pm):</td>
<td>0.06</td>
<td>20,657.48</td>
<td>327.96</td>
<td>161.64</td>
</tr>
</tbody>
</table>
### DSLOC Estimate Growth Model Version:

Version 8

### Version 8 DSLOC Estimate Growth Model Regression Method:

ODR

### DSLOC Estimate Growth Model Equations and Variables

#### New DSLOC Growth Equation:

\[
S_{\text{DGANew}} = \exp\left(- (\text{Decay} \times \text{Maturity}) \times (b_{[GN]} \times \epsilon_{[GU]} \times \left(\frac{S_{\text{DNew}}}{K_{[N]}}\right)^{a_{[GU]} \times K_{[N]} - S_{\text{DNew}}}\right) + S_{\text{DNew}}
\]

#### Modified DSLOC Growth Equation:

\[
S_{\text{DGAMod}} = \exp\left(- (\text{Decay} \times \text{Maturity}) \times (b_{[GM]} \times \epsilon_{[GM]} \times \left(\frac{S_{\text{DMod}}}{K_{[M]}}\right)^{a_{[GM]} \times K_{[M]} - S_{\text{DMod}}}\right) + S_{\text{DMod}}
\]

#### Unmodified DSLOC Growth Equation:

\[
S_{\text{DGAMod}} = \exp\left(- (\text{Decay} \times \text{Maturity}) \times (b_{[GU]} \times \epsilon_{[GU]} \times \left(\frac{S_{\text{DMod}}}{K_{[U]}}\right)^{a_{[GU]} \times K_{[U]} - S_{\text{DMod}}}\right) + S_{\text{DMod}}
\]

**where:**

\[
a_{[GN]} = 1.021, \quad a_{[GM]} = 0.913, \quad a_{[GU]} = 1.044
\]

\[
e = 2.7183, \quad \text{Decay} = 3.466
\]

### List Statistics

<table>
<thead>
<tr>
<th></th>
<th>[GN]</th>
<th>[GM]</th>
<th>[GU]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Data Points (observations):</td>
<td>225</td>
<td>136</td>
<td>142</td>
</tr>
<tr>
<td>Geometric (log space) mean of ( b ):</td>
<td>1.208E+00</td>
<td>2.651E+00</td>
<td>6.199E-01</td>
</tr>
<tr>
<td>Arithmetic (unit space) mean of ( b ):</td>
<td>1.736E+00</td>
<td>4.060E+00</td>
<td>7.510E-01</td>
</tr>
<tr>
<td>Standard deviation of ( b ):</td>
<td>1.849E+00</td>
<td>4.557E+00</td>
<td>6.566E-01</td>
</tr>
<tr>
<td>Coefficient of Variation (CV) ( b ):</td>
<td>1.07</td>
<td>1.12</td>
<td>0.87</td>
</tr>
<tr>
<td>Arithmetric (unit space) mean of ( \epsilon ):</td>
<td>1.366E+00</td>
<td>1.510E+00</td>
<td>1.191E+00</td>
</tr>
<tr>
<td>Standard deviation of ( \epsilon ):</td>
<td>1.224E+00</td>
<td>1.623E+00</td>
<td>9.296E-01</td>
</tr>
<tr>
<td>Coefficient of Variation (CV) of ( \epsilon ):</td>
<td>0.90</td>
<td>1.07</td>
<td>0.78</td>
</tr>
<tr>
<td>Mean Magnitude of the Relative Error:</td>
<td>44%</td>
<td>50%</td>
<td>24%</td>
</tr>
</tbody>
</table>

### New to Modified DSLOC Correlation:

2.570E-03

### New to Unmodified DSLOC Growth Correlation:

3.025E-01

### Growth Factor Estimating Relationships Behavior

<table>
<thead>
<tr>
<th></th>
<th>New DSLOC Growth</th>
<th>Modified DSLOC Growth</th>
<th>Unmodified DSLOC Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implied Growth Factor at data set mean baseline DSLOC:</td>
<td>53% at 59,443 DSLOC</td>
<td>11% at 22,934 DSLOC</td>
<td>7% at 251,323 DSLOC</td>
</tr>
<tr>
<td>Implied Growth Factor at data set geometric mean baseline DSLOC:</td>
<td>50% at 23,035 DSLOC</td>
<td>22% at 7,756 DSLOC</td>
<td>1% at 70,790 DSLOC</td>
</tr>
</tbody>
</table>
## Variables and Notation

### Variables Used in JCDER and DSLOC Estimate Growth Equations

\begin{align*}
S & \equiv \text{Size (ESLOC) distribution} \\
E & \equiv \text{Core Effort (person-months) or (pm) distribution (data set SDLC)} \\
T & \equiv \text{Core Duration (calendar months) or (cm) distribution (data set SDLC)} \\
E[A-A] & \equiv \text{All-All Effort (person-months) or (pm) distribution (All-All SDLC)} \\
T[A-A] & \equiv \text{All-All Duration (calendar months) or (cm) distribution (All-All SDLC)} \\
\alpha[E] & \equiv \text{Effort dimension exponent (economy/diseconomy of scale)} \\
\alpha[T] & \equiv \text{Duration dimension exponent (economy/diseconomy of scale)} \\
\alpha[S] & \equiv \text{Size dimension exponent (economy/diseconomy of scale)} \\
D & \equiv \text{Difficulty (nonlinear person-month months per ESLOC) distribution -- time-sensitive inverse productivity} \\
I & \equiv \text{Intensity (nonlinear person-months per calendar month) distribution -- staffing or burn} \\
\epsilon[GN] & \equiv \text{Baseline New DSLOC estimate growth error factor distribution} \\
\epsilon[GM] & \equiv \text{Baseline Modified DSLOC estimate growth error factor distribution} \\
\epsilon[GU] & \equiv \text{Baseline Unmodified DSLOC estimate growth error factor distribution} \\
Decay & \equiv \text{Decay Constant Parameter; default is 3.466 based on Boehm's "Cone of Uncertainty" (Boehm, 1981, p.311)} \\
\hat{b}[GN], \hat{b}[GM], \hat{b}[GU] & \equiv \text{Geometric mean scale factor parameters for New, Modified, and Unmodified DSLOC growth estimating relationships} \\
Maturity & \equiv \text{Estimate Maturity Parameter (e.g., SDLBegin=0%; SyRR=10%; SWRR=20%; SwPDR=40%; SWCDR=60%; TRR=80%; SWAccept=100%)} \\
K[N], K[M], K[U] & \equiv \text{Software Item (SI) to CSCI normalization factor for New, Mod, and Umod DSLOC} \\
S[DN] & \equiv \text{Technical Baseline Estimate (TBE) of New DSLOC} \\
S[DM] & \equiv \text{Technical Baseline Estimate (TBE) of Modified DSLOC} \\
S[DU] & \equiv \text{Technical Baseline Estimate (TBE) of Unmodified DSLOC} \\
S[DGN] & \equiv \text{Growth-adjusted New DSLOC estimate distribution} \\
S[DGM] & \equiv \text{Growth-adjusted Modified DSLOC estimate distribution} \\
S[DGU] & \equiv \text{Growth-adjusted Unmodified DSLOC estimate distribution}
\end{align*}

### Notation Conventions

- ‘≡’ is used to indicate identity; left expression is defined by right expression.
- ‘≜’ is used to indicate estimation; the right expression estimates the left expression.

- **Arial Bold Italic** font indicates a random variable; range and distribution of possible outcomes.

- **Times New Roman Italic** font indicates a specific-value (single-value) variable.

- Bracks ‘[ ]’ are used to indicate the subscripted characters of a variable name.
Appendix C       DEGM8 vs. DEGM7 Implied Growth Factor Comparison

This appendix contains three scatter charts that show the implied growth percentage for each of New, Modified, and Unmodified DSLOC as a function of the given TBE DSLOC. We can summarize the information on these charts as follows:

- New DSLOC Implied Growth Percentage at SDLCBegin (Maturity = 0%) for TBE DSLOC in the range 137 DSLOC to 858,984 DSLOC (225 observations)
  - DEGM8 Default: 34% to 62%
  - DEGM7: 75%

- Modified DSLOC Implied Growth Percentage at SDLCBegin (Maturity = 0%) for TBE DSLOC in the range 10 DSLOC to 208,082 DSLOC (136 observations)
  - DEGM8 Default: 117% to -8%
  - DEGM7: 43%

- Unmodified DSLOC Implied Growth Percentage at SDLCBegin (Maturity = 0%) for TBE in the range 1,100 DSLOC to 6,564,104 DSLOC (142 observations)
  - DEGM8 Default: -16% to 24%
  - DEGM7: 43%
**Figure 10** Filtered 2015 SRDR New DSLOC data regressed with ODR in log space (DEGM8 default) yields a better fit than does simply using the DEGM7 relationship.

**Figure 11** Filtered 2015 SRDR Modified DSLOC data regressed with ODR in log space (DEGM8 default) yields a better fit than does simply using the DEGM7 relationship.
Figure 12  Filtered 2015 SRDR Unmodified DSLOC data regressed with ODR in log space (DEGM8 default) yields a better fit than does simply using the DEGM7 relationship.
Appendix D  

Orthogonal Distance Regression

Orthogonal Distance Regression (ODR) is the name given to the computational problem associated with finding the maximum likelihood estimators of parameters in measurement error models in the case of normally distributed errors. For some data set \( P \) that contains \( n \) observations, each containing unique values for the same \( m \) observable parameters, ODR seeks to find a line in \( \mathbb{R}^m \) space that minimizes the sum of the squared orthogonal (shortest) distances between each data point and that line. In other words, ODR is a process for finding a best fit line (an estimator) through a multi-dimension set of data points (observations).

Why is ODR better than Ordinary Least Squares (OLS) regression and its variants?

- ODR works in situations where there are more than two dimensions (measures) without making assumptions about which measures are dependent and which are independent.
- ODR acknowledges the existence of measurement error in all dimensions; not just in the dimension associated with a single variable deemed to be “dependent”.

Data Sets

Let \( P \) be a set of data points (relevant measure observations) in \( \mathbb{R}^m \) space. We can describe \( P \) as an \( n \times m \) matrix of data point coordinate values

\[
P_{i,j} \in P \equiv \begin{bmatrix}
P_{1,1} & P_{1,2} & \cdots & P_{1,m} \\
P_{2,1} & P_{2,2} & \cdots & P_{2,m} \\
\vdots & \vdots & \ddots & \vdots \\
P_{n,1} & P_{n,2} & \cdots & P_{n,m}
\end{bmatrix}
\]  

(90)

where \( m \) is the number of dimensions (relevant measures) and where \( n \) is the number of observations (data points).

Data Set Centroid

Given the data set \( P \) of \( m \) relevant measures and \( n \) observations, we define the centroid point \( C_p \) of \( P \) with position vectors \( r_i \) as
Definition: Data Set Centroid

\[ C_p = \frac{1}{n} \sum_{i=1}^{n} r_i \quad | P_{i,j} \in P \subseteq \mathbb{R}^m \]

\[ = \left( \frac{1}{n} \sum_{i=1}^{n} P_{i,1}, \frac{1}{n} \sum_{i=1}^{n} P_{i,2}, \ldots, \frac{1}{n} \sum_{i=1}^{n} P_{i,m} \right) \]  

(91)

Orthogonal Distance Regression to Find the Best Fit Line

Orthogonal Distance Regression (ODR) (a special case of Total Least Squares regression) in m-dimension space (\( \mathbb{R}^m \)) can be used to determine the equation of a best fit line \( L_{ODR} \) according to the following definition:

Definition: ODR Best Fit Line

An ODR best fit line \( L_{ODR} \) in \( \mathbb{R}^m \) space is one that minimizes the sum of the squared orthogonal distances \( \sum_{i=1}^{n} \delta_i^2 \) from \( L_{ODR} \) to each point \( P_i \) of a given set of data points \( P \) in \( \mathbb{R}^m \).

One way to specify the best fit line \( L_{ODR} \) is to specify a particular point \( P' \) on \( L_{ODR} \) and to specify a direction vector \( a \); i.e., a vector that is collinear with or parallel to \( L_{ODR} \).

Specifying a Point on the ODR Best Fit Line

We first specify an arbitrary reference point on \( L_{ODR} \) which we call \( P' \) with position vector \( r' \). We next identify each of points \( \tilde{P}_i \) on \( L_{ODR} \) with associated position vectors \( \tilde{r}_i \) from which we measure the orthogonal distance to each corresponding data point \( P_i \) with position vector \( r_i \). In this appendix we use the overstrike tilde (\( \tilde{\cdot} \)) notation on a point or a position vector to indicate an estimated value. In this case \( \tilde{P}_i \) is an estimate of its corresponding actual data point \( P_i \) since, practically speaking, if \( L_{ODR} \) represents the best fit line and \( \tilde{P}_i \) is, by definition, on \( L_{ODR} \), then \( \tilde{P}_i \) represents the best estimate of actual outcome \( P_i \). This implies each vector \( \overrightarrow{P_i\tilde{P}_i} \) is orthogonal to \( L_{ODR} \) and, hence, any one of these vectors qualifies as a normal vector of \( L_{ODR} \). Since each vector \( \overrightarrow{P_i\tilde{P}_i} \) is obviously on \( L_{ODR} \), it follows that each \( \overrightarrow{P_i\tilde{P}_i} \) is orthogonal to each \( \overrightarrow{\tilde{P}_iP_i} \).
Figure 13 Geometry of the orthogonal distance problem; maximizing the sum of the squared angles $\theta_i$ serves to minimize the sum of the squared orthogonal distances $\|n_i\|

The previous paragraph and Figure 13 describe a right triangle in $\mathbb{R}^m$ space with base leg $a_i \equiv P_iP'_{i}$ (on $L_{ODR}$), normal leg $n_i \equiv P_iP'_{i}$ (orthogonal to $L_{ODR}$), and hypotenuse $v_i \equiv P'_{i}$. By the angle between vectors definition

**Definition: Angle Between Vectors**

$$
\cos \theta = \frac{u \cdot v}{\|u\| \|v\|}
$$

$$
= \frac{\sum_{i=1}^{m} u_i v_i}{\sqrt{\sum_{i=1}^{m} u_i^2} \sqrt{\sum_{i=1}^{m} v_i^2}}
$$

(92)

and the dot product associativity theorem,

**Theorem: Dot Product Associativity**

$$(cu) \cdot v = c(u \cdot v)$$

(93)
we define the angle $\theta_i$ between the unit vector of any normal leg $\hat{n}$ and each hypotenuse vector $v_i$ to be

$$
\theta_i \mid \cos \theta_i = \frac{\hat{n} \cdot v_i}{\|\hat{n}\| \|v_i\|} = \frac{1}{\|v_i\|} (\hat{n} \cdot v_i)
$$

(94)

Substituting vector $v_i$ in Equation (94) with its equivalent position vector difference $r_i - r'$ per the definition of a direction vector

**Definition: Direction Vector**

$$
v = k (r'' - r') \quad v, r', r'' \in \mathbb{R}^m; k \in \mathbb{R}
$$

(95)

gives us

$$
\hat{n} \cdot (r_i - r') = \|r_i - r'\| \cos \theta_i
$$

(96)

By the dot product distributivity theorem

**Theorem: Dot Product Distributivity**

$$
u \cdot (v + w) = u \cdot v + u \cdot w
$$

(97)

we can rewrite Equation (96) as

$$
\hat{n} \cdot r_i - \hat{n} \cdot r' = \|r_i - r'\| \cos \theta_i
$$

(98)

If we let $d \equiv \hat{n} \cdot r_i - \|r_i - r'\| \cos \theta_i$ then we can say

$$
d = \hat{n} \cdot r'
$$

(99)

From the geometry of Figure 13 above, it should be obvious that minimizing the sum of the squared magnitudes of each vector $n_i$ can be accomplished by maximizing the sum of each squared $\theta_i$ which, by inspection of Equations (94) and (99), implies minimizing the sum of each squared $d_i$. We have already specified the locations of points $P_i$ (the given data points) and point $P'$ (the arbitrary reference point on $L_{ODR}$) which necessarily specifies the magnitudes of vectors $v_i$. Because each vector $v_i$ is the
hypotenuse of a right triangle, the magnitude of each normal leg \( \|n_i\| \) can be determined as the magnitude of the vector projection of each \( v_i \) onto the unit vector \( \hat{n} \). The unit vector \( \hat{n} \), per the vector unitizing theorem, 

**Theorem: Unitizing a Vector**

\[
\hat{u} = \frac{1}{\|v\|} v
\]  

(100)

is equal to \( n_i/\|n_i\| \). The magnitude \( \|n_i\| \) of each vector \( n_i \) can therefore be specified as

**Definition: Normal Vector Magnitude**

\[
\|n_i\| = \left\| proj_{\hat{n}_i} (v_i) \right\| \quad n_i \equiv \frac{n_i}{\|n_i\|}
\]  

(101)

By the vector projection definition

**Definition: Vector Projection**

\[
proj_u (v) = \left( \frac{u \cdot v}{u \cdot u} \right) u
\]  

(102)

we can rewrite Definition (101) as

\[
\|n_i\| = \left\| \frac{\hat{n} \cdot v_i}{\hat{n} \cdot \hat{n}} \hat{n} \right\|
\]  

(103)

Substituting vector \( v_i \) in Equation (103) with its equivalent position vector difference \( r_i - r' \) per Definition (95) gives us

\[
\|n_i\| = \left\| \frac{\hat{n} \cdot (r_i - r')}{\hat{n} \cdot \hat{n}} \hat{n} \right\|
\]  

(104)

By the dot product associativity Theorem (97), Equation (104) becomes

\[
\|n_i\| = \left\| \frac{\hat{n} \cdot r_i - \hat{n} \cdot r'}{\hat{n} \cdot \hat{n}} \hat{n} \right\|
\]  

(105)
We next substitute $\hat{n} \cdot \hat{n}$ in Equation (105) with its equivalent $\|\hat{n}\|^2$ implied by the definition of vector magnitude

**Definition: Vector Magnitude**

$$\|v\| = \sqrt{v \cdot v} = \sqrt{v_1^2 + v_2^2 + \cdots + v_m^2}$$

which yields

$$\|n_i\| = \left\| \frac{\hat{n} \cdot r_i - \hat{n} \cdot r'}{\|\hat{n}\|^2} \right\| (\hat{n})$$

(107)

Observing that the quantities $\hat{n} \cdot r_i$, $\hat{n} \cdot r'$, and $\|\hat{n}\|^2$ all evaluate to scalar values, we can use the vector magnitude factoring theorem

**Theorem: Vector Magnitude Factoring**

$$\|cv\| = c\|v\|$$

(108)

to rewrite Equation (107) as

$$\|n_i\| = \frac{\hat{n} \cdot r_i - \hat{n} \cdot r'}{\|\hat{n}\|^2} \|\hat{n}\|$$

$$= \frac{\hat{n} \cdot r_i - \hat{n} \cdot r'}{\|\hat{n}\|}$$

$$= \hat{n} \cdot r_i - \hat{n} \cdot r'$$

(109)

Substituting $\hat{n} \cdot r'$ in Equation (109) with its equivalent $d$ in Equation (99) yields

$$\|n_i\| = \hat{n} \cdot r_i \quad d$$

(110)

The sum of the squared normal vector magnitudes can now be written as
\[
\sum_{i=1}^{n} \|\mathbf{n}_i\|^2 = \sum_{i=1}^{n} (\hat{\mathbf{n}} \cdot \mathbf{r}_i - d)^2
\]  \hspace{1cm} (111)

We define a function \( f \) to represent our sum of squared vector magnitudes

\[
f(\hat{\mathbf{n}}, d | \mathbf{r}) = \sum_{i=1}^{n} \|\mathbf{n}_i\|^2 = \sum_{i=1}^{n} (\hat{\mathbf{n}} \cdot \mathbf{r}_i - d)^2
\]  \hspace{1cm} (112)

We can solve for the value of each \( d \) that minimizes \( f \) by setting the partial derivative of \( f \) with respect to \( d \) in Equation (112) to zero. This yields

\[
\frac{\partial}{\partial d} f = \frac{\partial}{\partial d} \sum_{i=1}^{n} (\hat{\mathbf{n}} \cdot \mathbf{r}_i - d)^2 = 0
\]  \hspace{1cm} (113)

To solve for \( d \) we first expand the square quantity within the summation.

\[
\frac{\partial}{\partial d} \sum_{i=1}^{n} \left( (\hat{\mathbf{n}} \cdot \mathbf{r}_i)^2 - 2(\hat{\mathbf{n}} \cdot \mathbf{r}_i) d + d^2 \right) = 0
\]  \hspace{1cm} (114)

Next we apply the summation to each term within the summation and factor out the scalars.

\[
\frac{\partial}{\partial d} \left( \sum_{i=1}^{n} (\hat{\mathbf{n}} \cdot \mathbf{r}_i)^2 - 2d \sum_{i=1}^{n} (\hat{\mathbf{n}} \cdot \mathbf{r}_i) + d^2 n \right) = 0
\]  \hspace{1cm} (115)

Taking the partial derivative of each term with respect to \( d \) gives us

\[
\frac{\partial}{\partial d} \left( \sum_{i=1}^{n} (\hat{\mathbf{n}} \cdot \mathbf{r}_i)^2 \right) - \frac{\partial}{\partial d} \left( 2d \sum_{i=1}^{n} (\hat{\mathbf{n}} \cdot \mathbf{r}_i) \right) + \frac{\partial}{\partial d} (d^2 n) = 0
\]  \hspace{1cm} (116)

Resolving the partial derivatives yields

\[
\frac{\partial}{\partial d} \left( \sum_{i=1}^{n} (\hat{\mathbf{n}} \cdot \mathbf{r}_i)^2 \right) - \frac{\partial}{\partial d} \left( 2d \sum_{i=1}^{n} (\hat{\mathbf{n}} \cdot \mathbf{r}_i) \right) + \frac{\partial}{\partial d} (d^2 n) = 0
\]  \hspace{1cm} (117)

\[
\therefore -2 \sum_{i=1}^{n} \hat{\mathbf{n}} \cdot \mathbf{r}_i + 2dn = 0
\]
Since \( \hat{n} \) is a constant vector (i.e., is the same for each \( r_i \)), we can use the dot product distributivity Theorem (97) to remove \( \hat{n} \) from the summation to get

\[
-2\hat{n} \cdot \sum_{i=1}^{n} r_i + 2dn = 0
\]  

(118)

Multiplying the summation by \( \frac{n}{n} \) (the equivalent of 1) gives us

\[
-2\hat{n} \cdot \frac{n}{n} \sum_{i=1}^{n} r_i + 2dn = 0
\]

(119)

\[
\therefore -2\hat{n} \cdot \frac{1}{n} \sum_{i=1}^{n} r_i + 2dn = 0
\]

We observe that Equation (119) contains an expression \((1/n)\sum_{i=1}^{n} r_i\) that is equivalent to the data set centroid \( C_p \) in Definition(91). Making the substitution yields

\[
-2\hat{n} \cdot nC_p + 2dn = 0
\]  

(120)

We now apply dot product commutativity

**Theorem: Dot Product Commutativity**

\[
u \cdot v = v \cdot u
\]  

(121)

and dot product associativity Theorem (93) to Equation(120), the result being

\[
-2n(\hat{n} \cdot C_p) + 2dn = 0
\]  

(122)

Finally, solving for \( d \) gives us

\[
-2n(\hat{n} \cdot C_p) + 2dn = 0
\]

\[
\therefore d = \hat{n} \cdot C_p
\]  

(123)

Comparing Equation (123) to Equation (99) leads to the conclusion that the arbitrary reference point \( P' \) on \( L_{ODR} \) and with position vector \( r' \) is equal to the data set centroid \( C_p \) which proves
Theorem: The ODR Best Fit Line Contains the Data Set Centroid

\[ C_P \in L_{ODR} \quad (124) \]

Once again, a way to specify \( L_{ODR} \) is to specify a particular point \( P' \) on the line and to specify a direction vector \( \vec{a} \); i.e., a vector that is collinear with or parallel to the line. Substituting the appropriate vector components for \( \vec{a} \), \( \vec{r}_i \), and \( C_P \) into the generalized vector equation of a line in multi-dimension space

**Theorem: Vector Equation of a Line in \( m \)-Dimension Space**

\[
\vec{r} = \vec{r}' + t \vec{a} \text{ for } \vec{r}, \vec{r}', \vec{a} \in \mathbb{R}^m 
\quad (125)
\]
gives us

**Theorem: The ODR Best Fit Line, Vector Form**

\[
\vec{r} = C_P + t \vec{a} \Rightarrow \vec{r} \in L_{ODR} \text{ for } \vec{r}, C_P, \vec{a} \in \mathbb{R}^m 
\quad (126)
\]

Substituting the vector component form

**Definition: Vector Components and Notations**

\[
v = [v_1, v_2, \ldots, v_m] \\
v = \begin{bmatrix} v_1 & v_2 & \cdots & v_m \end{bmatrix} \\
v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{bmatrix} 
\quad (127)
\]
of vectors \( \vec{r} \), \( C_P \), and \( \vec{a} \) into the equation of Theorem (126) gives us
The parametric form equation of $L_{ODR}$ is therefore

**Theorem: ODR Best Fit Line, Parametric Form**

$$\begin{align*}
\tilde{r}_1 &= C_{P1} + t a_1 \\
\tilde{r}_2 &= C_{P2} + t a_2 \\
&\vdots \\
\tilde{r}_m &= C_{Pm} + t a_m 
\end{align*}$$

(129)

The scalar or implicit form equation of $L_{ODR}$ is therefore

**Theorem: ODR Best Fit Line, Symmetric Form**

$$\begin{align*}
\left(\frac{\tilde{r}_1 - C_{P1}}{a_1}\right) &= \left(\frac{\tilde{r}_2 - C_{P2}}{a_2}\right) = \cdots = \left(\frac{\tilde{r}_m - C_{Pm}}{a_m}\right) 
\end{align*}$$

(130)

**Using Singular Value Decomposition to Specify a Direction Vector for the ODR Best Fit Line**

Since we can easily specify a particular point on the ODR best fit line $L_{ODR}$ by computing the centroid $C_p$ from the values in our data set $P$ using Definition (91)

$$C_p = \frac{1}{n} \sum_{i=1}^{n} r_i$$

(131)

we need only find values for the direction vector $[a_1, a_2, \ldots, a_m]$ in order to fully specify $L_{ODR}$.

We can find these values by using Principal Component Analysis (PCA) to perform a principal component transformation on our data set matrix $P$. PCA is mathematically defined as an orthogonal linear transformation that transforms the data to a new coor-
ordinate system such that the greatest variance by some projection of the data comes to lie
on the first coordinate (called the first principal component), the second greatest vari-
ance on the second coordinate, and so on. It turns out that the first principal compo-
nent axis has the same direction as does $L_{ODR}$. One way to perform a principal compo-
nent transformation is to use the *Singular Value Decomposition* (SVD). The SVD is a
special matrix factorization that, when applied to a data-set-centroid-centered version
our data set matrix $P$, will yield, among other things, the components of the $L_{ODR}$ direc-
tion vector $[a_1 \ a_2 \ \cdots \ a_m]$. To apply SVD, we first center our data set matrix $P$ about the centroid
point $C_P$ to create an $n \times m$ matrix $M$; i.e., we create $M$ as the differences in each di-
mension between each data point $P_i$ and the data set centroid $C_P$ such that

$$M = \begin{bmatrix}
P_{1,1} - C_{P_1} & P_{1,2} - C_{P_2} & \cdots & P_{1,m} - C_{P_m} \\
P_{2,1} - C_{P_1} & P_{2,2} - C_{P_2} & \cdots & P_{2,m} - C_{P_m} \\
\vdots & \vdots & \ddots & \vdots \\
P_{n,1} - C_{P_1} & P_{n,2} - C_{P_2} & \cdots & P_{n,m} - C_{P_m}
\end{bmatrix}$$

(132)

The SVD of matrix $M$ is a factorization of $M$ such that

$$SVD(M) = \{U, \Sigma, V^T\} \quad | M = U\Sigma V^T$$

(133)

where

- $M$ = Given centroid-centered data set matrix $(n \times m)$
- $U$ = Orthogonal matrix $(n \times p)$; $p = \min(n, m)$; not used as part of
  the ODR process
- $\Sigma$ = Square diagonal matrix $(p \times p)$; singular values
- $V$ = Orthogonal matrix $(m \times p)$; singular vectors organized by
  row; $V^T$ contains singular vectors organized by column
Note that a feature of the algorithm being used to implement SVD is that it returns $V$ such that its contained vectors (columns) are sorted in descending singular value order from left to right. PCA asserts that the ODR best fit line’s direction vector $a$ is equal to the direction of PCA’s first principal component axis which is the singular vector of $M$ that corresponds to its largest singular value; in this implementation, the leftmost column vector in $V$. Therefore

$$
\begin{bmatrix}
a_1 \\
a_2 \\
\vdots \\
a_m
\end{bmatrix}
= 
\begin{bmatrix}
V_{1,1} \\
V_{2,1} \\
\vdots \\
V_{3,1}
\end{bmatrix}
$$

(135)

In summary, we can find the system of equations that define an ODR best fit line through any multi-dimension commensurable and scale-invariant data set by first finding the data set centroid, then centering the data set matrix about the data space origin, and finally applying the SVD to the centered matrix. The resulting ODR best fit line is specified by a known point on the line (the data set centroid) and a direction vector (the column of the singular vector matrix that is associated with the largest singular value in the singular value matrix).
Appendix E  

TRI ACEIT Implementation of the Example Estimate

The following two tables show results from the DEGM8 example estimate described in the body of this paper. They compare DEGM8 idealized values produced by an MS Excel workbook against the output of an implementation of the DEGM8 example estimate in the TRI ACEIT tool.

Table 4  Example estimate DSLOC values comparison; MS Excel idealized from data set versus ACEIT risk output; differences in Total DSLOC because MS Excel assumes no correlation between summands (independence) while ACEIT uses given DEGM8 correlation values (group strength) between each summand in a Monte Carlo convolution (partial dependence). Notice that the Total DSLOC mean values are very close to each other which is expected since probability theory proves that the sum of the means is equal to the mean of the sum.

<table>
<thead>
<tr>
<th></th>
<th>TBE DSLOC</th>
<th>MS Excel Geomean DSLOC</th>
<th>ACEIT Estimate DSLOC</th>
<th>MS Excel Median Position DSLOC</th>
<th>ACEIT Median DSLOC</th>
<th>MS Excel Mean DSLOC</th>
<th>ACEIT Mean DSLOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total DSLOC</td>
<td>175,000</td>
<td>183,600</td>
<td>183,703</td>
<td>190,325</td>
<td>192,755</td>
<td>213,666</td>
<td>213,672</td>
</tr>
<tr>
<td>Attainment Probability</td>
<td>#N/A</td>
<td>#N/A</td>
<td>30%</td>
<td>#N/A</td>
<td>50%</td>
<td>#N/A</td>
<td>68%</td>
</tr>
<tr>
<td>Implied Growth</td>
<td>0%</td>
<td>5%</td>
<td>5%</td>
<td>9%</td>
<td>10%</td>
<td>22%</td>
<td>22%</td>
</tr>
<tr>
<td>New DSLOC</td>
<td>25,000</td>
<td>31,217</td>
<td>31,227</td>
<td>33,769</td>
<td>33,768</td>
<td>38,091</td>
<td>38,090</td>
</tr>
<tr>
<td>Attainment Probability</td>
<td>28%</td>
<td>41%</td>
<td>41%</td>
<td>50%</td>
<td>50%</td>
<td>62%</td>
<td>62%</td>
</tr>
<tr>
<td>Implied Growth</td>
<td>0%</td>
<td>25%</td>
<td>25%</td>
<td>35%</td>
<td>35%</td>
<td>52%</td>
<td>52%</td>
</tr>
<tr>
<td>Total DSLOC</td>
<td>50,000</td>
<td>50,965</td>
<td>50,970</td>
<td>54,342</td>
<td>54,344</td>
<td>64,226</td>
<td>64,227</td>
</tr>
<tr>
<td>Attainment Probability</td>
<td>39%</td>
<td>43%</td>
<td>43%</td>
<td>50%</td>
<td>50%</td>
<td>75%</td>
<td>75%</td>
</tr>
<tr>
<td>Implied Growth</td>
<td>0%</td>
<td>2%</td>
<td>2%</td>
<td>9%</td>
<td>9%</td>
<td>28%</td>
<td>28%</td>
</tr>
<tr>
<td>Unmodified DSLOC</td>
<td>100,000</td>
<td>101,418</td>
<td>101,506</td>
<td>102,213</td>
<td>102,213</td>
<td>111,350</td>
<td>111,355</td>
</tr>
<tr>
<td>Attainment Probability</td>
<td>39%</td>
<td>46%</td>
<td>47%</td>
<td>50%</td>
<td>50%</td>
<td>77%</td>
<td>76%</td>
</tr>
<tr>
<td>Implied Growth</td>
<td>0%</td>
<td>1%</td>
<td>2%</td>
<td>2%</td>
<td>2%</td>
<td>11%</td>
<td>11%</td>
</tr>
</tbody>
</table>
Table 5 Example estimate CDFs comparison; MS Excel idealized from data set versus ACEIT risk output; differences in Total DSLOC table because MS Excel assumes no correlation between summands (independence) while ACEIT uses given DEGM8 correlation values (group strength) between each summand in a Monte Carlo convolution (partial dependence).

<table>
<thead>
<tr>
<th>%ile</th>
<th>Total DSLOC</th>
<th>New DSLOC</th>
<th>Modified DSLOC</th>
<th>Unmodified DSLOC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MS Excel CDF (DSLOC) (no correl)</td>
<td>ACE CDF (DSLOC) (w/ correl)</td>
<td>%ile</td>
<td>MS Excel CDF (DSLOC)</td>
</tr>
<tr>
<td>5</td>
<td>113,123</td>
<td>144,611</td>
<td>5</td>
<td>16,767</td>
</tr>
<tr>
<td>10</td>
<td>126,853</td>
<td>154,073</td>
<td>10</td>
<td>18,002</td>
</tr>
<tr>
<td>15</td>
<td>140,940</td>
<td>160,765</td>
<td>15</td>
<td>20,367</td>
</tr>
<tr>
<td>20</td>
<td>149,399</td>
<td>166,549</td>
<td>20</td>
<td>22,064</td>
</tr>
<tr>
<td>30</td>
<td>168,602</td>
<td>175,190</td>
<td>30</td>
<td>26,077</td>
</tr>
<tr>
<td>35</td>
<td>176,094</td>
<td>179,519</td>
<td>35</td>
<td>28,673</td>
</tr>
<tr>
<td>40</td>
<td>180,591</td>
<td>184,028</td>
<td>40</td>
<td>30,255</td>
</tr>
<tr>
<td>45</td>
<td>186,030</td>
<td>188,374</td>
<td>45</td>
<td>32,652</td>
</tr>
<tr>
<td>50</td>
<td>190,325</td>
<td>192,755</td>
<td>50</td>
<td>33,769</td>
</tr>
<tr>
<td>55</td>
<td>193,776</td>
<td>197,488</td>
<td>55</td>
<td>35,318</td>
</tr>
<tr>
<td>60</td>
<td>197,515</td>
<td>202,922</td>
<td>60</td>
<td>37,473</td>
</tr>
<tr>
<td>65</td>
<td>204,400</td>
<td>209,970</td>
<td>65</td>
<td>39,469</td>
</tr>
<tr>
<td>70</td>
<td>208,437</td>
<td>213,239</td>
<td>70</td>
<td>40,242</td>
</tr>
<tr>
<td>75</td>
<td>215,360</td>
<td>230,280</td>
<td>75</td>
<td>40,882</td>
</tr>
<tr>
<td>80</td>
<td>226,514</td>
<td>246,547</td>
<td>80</td>
<td>43,097</td>
</tr>
<tr>
<td>85</td>
<td>250,747</td>
<td>272,034</td>
<td>85</td>
<td>47,510</td>
</tr>
<tr>
<td>90</td>
<td>333,950</td>
<td>307,845</td>
<td>90</td>
<td>62,711</td>
</tr>
<tr>
<td>95</td>
<td>434,026</td>
<td>364,916</td>
<td>95</td>
<td>84,782</td>
</tr>
</tbody>
</table>

The remaining tables in this appendix illustrate an implementation of the DEGM8 in the TRI ACEIT software tool that models the example estimate. These tables are listed in the order in which they must appear in the AECIT file. The order of computation is from last appearing to first appearing; each table referencing necessary predecessor tables through relative (indexed) addressing supported by the index values in the Ri and i columns of each referencing table. The Equation / Throughput column holds ACEIT expressions that represent various parts of the DEGM8 equations.
Table 6  Growth Adjusted Accumulated DSLOC Distributions Table

<table>
<thead>
<tr>
<th>WBS/CES Description</th>
<th>Unique ID</th>
<th>Equation / Throughput</th>
<th>WBS Indent Level</th>
<th>PE Position in Distribution</th>
<th>Distribution Form</th>
<th>Grouping</th>
<th>Group Strength</th>
<th>Random Seed</th>
<th>CDF Keyword</th>
<th>Ri (+) Row Index</th>
<th>i (+) Row Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>*EXAMPLE - DSLOC Estimate Growth Model v8 (DEGM8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* Growth Adjusted Accumulated DSLOC Distributions (includes Redelivery) (combined)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* THIS TABLE MUST BE IN REVERSE ORDER</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Table Header Row -- DO NOT DELETE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(table rows must be at WBS Level 2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example Software Item with Unmodified DSLOC Only</td>
<td></td>
<td>FYIVal[@GMSdNTbl+i]+FYIVal[@GMSdNTbl+i]+FYIVal[@GMSdNTbl+i]+FYIVal[@GMSdNTbl+i]+FYIVal[@GMSdNTbl+i]+FYIVal[@GMSdNTbl+i]+FYIVal[@GMSdNTbl+i]+FYIVal[@GMSdNTbl+i]+FYIVal[@GMSdNTbl+i]</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example Software Item with Modified DSLOC Only</td>
<td></td>
<td>exp(-3.466<em>FYIVal[@STbl+i,2012])</em>(FYTot(@GSdNTbl+i)-FYIVal[@STbl+i,2011])+FYIVal[@STbl+i,2011]</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example Software Item with New DSLOC Only</td>
<td></td>
<td>exp(-3.466<em>FYIVal[@STbl+i,2012])</em>(FYTot(@GSdNTbl+i)-FYIVal[@STbl+i,2011])+FYIVal[@STbl+i,2011]</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example Software Item with New, Modified, and Unmodified DSLOC</td>
<td></td>
<td>exp(-3.466<em>FYIVal[@STbl+i,2012])</em>(FYTot(@GSdNTbl+i)-FYIVal[@STbl+i,2011])+FYIVal[@STbl+i,2011]</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7  Growth/Maturity-Adjusted New DSLOC Distributions Table

<table>
<thead>
<tr>
<th>WBS/CES Description</th>
<th>Unique ID</th>
<th>Equation / Throughput</th>
<th>WBS Indent Level</th>
<th>PE Position in Distribution</th>
<th>Distribution Form</th>
<th>Grouping</th>
<th>Group Strength</th>
<th>Random Seed</th>
<th>CDF Keyword</th>
<th>Ri (+) Row Index</th>
<th>i (+) Row Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>* Growth/Maturity-Adjusted New DSLOC Distributions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Table Header Row -- DO NOT DELETE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(table rows must be at WBS Level 2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example Software Item with New, Modified, and Unmodified DSLOC</td>
<td></td>
<td>exp(-3.466<em>FYIVal[@STbl+i,2012])</em>(FYTot(@GSdNTbl+i)-FYIVal[@STbl+i,2011])+FYIVal[@STbl+i,2011]</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example Software Item with New DSLOC Only</td>
<td></td>
<td>exp(-3.466<em>FYIVal[@STbl+i,2012])</em>(FYTot(@GSdNTbl+i)-FYIVal[@STbl+i,2011])+FYIVal[@STbl+i,2011]</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example Software Item with Modified DSLOC Only</td>
<td></td>
<td>exp(-3.466<em>FYIVal[@STbl+i,2012])</em>(FYTot(@GSdNTbl+i)-FYIVal[@STbl+i,2011])+FYIVal[@STbl+i,2011]</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example Software Item with Unmodified DSLOC Only</td>
<td></td>
<td>exp(-3.466<em>FYIVal[@STbl+i,2012])</em>(FYTot(@GSdNTbl+i)-FYIVal[@STbl+i,2011])+FYIVal[@STbl+i,2011]</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 8 Growth/Maturity-Adjusted Modified DSLOC Distributions Table

<table>
<thead>
<tr>
<th>WBS/CES Description</th>
<th>Unique ID</th>
<th>Equation / Throughput</th>
<th>WBS Indent Level</th>
<th>PE Position in Distribution</th>
<th>Grouping</th>
<th>Group Strength</th>
<th>Random Seed</th>
<th>CDF Keyword</th>
<th>Ri (+) Rev Row Index</th>
<th>i (+) Row Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example Software Item with New, Modified, and Unmodified DSLOC Only</td>
<td><a href="#">GMSdMTbl</a></td>
<td>$\exp(-3.466 \times FYIVal(@STbl+i,2015)) \times (FYTot(@GSdMTbl+i) - (FYIVal(@STbl+i,2013) - FYIVal(@STbl+i,2014))) + (FYIVal(@STbl+i,2013) - FYIVal(@STbl+i,2014))$ [DEGM8]</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example Software Item with New DSLOC Only</td>
<td><a href="#">GMSdMTbl</a></td>
<td>$\exp(-3.466 \times FYIVal(@STbl+i,2015)) \times (FYTot(@GSdMTbl+i) - (FYIVal(@STbl+i,2013) - FYIVal(@STbl+i,2014))) + (FYIVal(@STbl+i,2013) - FYIVal(@STbl+i,2014))$ [DEGM8]</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Example Software Item with Modified DSLOC Only</td>
<td><a href="#">GMSdMTbl</a></td>
<td>$\exp(-3.466 \times FYIVal(@STbl+i,2015)) \times (FYTot(@GSdMTbl+i) - (FYIVal(@STbl+i,2013) - FYIVal(@STbl+i,2014))) + (FYIVal(@STbl+i,2013) - FYIVal(@STbl+i,2014))$ [DEGM8]</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Example Software Item with Unmodified DSLOC Only</td>
<td><a href="#">GMSdMTbl</a></td>
<td>$\exp(-3.466 \times FYIVal(@STbl+i,2015)) \times (FYTot(@GSdMTbl+i) - (FYIVal(@STbl+i,2013) - FYIVal(@STbl+i,2014))) + (FYIVal(@STbl+i,2013) - FYIVal(@STbl+i,2014))$ [DEGM8]</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

### Table 9 Growth/Maturity-Adjusted Unmodified DSLOC Distributions Table

<table>
<thead>
<tr>
<th>WBS/CES Description</th>
<th>Unique ID</th>
<th>Equation / Throughput</th>
<th>WBS Indent Level</th>
<th>PE Position in Distribution</th>
<th>Grouping</th>
<th>Group Strength</th>
<th>Random Seed</th>
<th>CDF Keyword</th>
<th>Ri (+) Rev Row Index</th>
<th>i (+) Row Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example Software Item with New, Modified, and Unmodified DSLOC Only</td>
<td><a href="#">GMSdUTbl</a></td>
<td>$\exp(-3.466 \times FYIVal(@STbl+i,2022)) \times (FYTot(@GSdUTbl+i) - (FYIVal(@STbl+i,2020) - FYIVal(@STbl+i,2021))) + (FYIVal(@STbl+i,2020) - FYIVal(@STbl+i,2021))$ [DEGM8]</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example Software Item with New DSLOC Only</td>
<td><a href="#">GMSdUTbl</a></td>
<td>$\exp(-3.466 \times FYIVal(@STbl+i,2022)) \times (FYTot(@GSdUTbl+i) - (FYIVal(@STbl+i,2020) - FYIVal(@STbl+i,2021))) + (FYIVal(@STbl+i,2020) - FYIVal(@STbl+i,2021))$ [DEGM8]</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Example Software Item with Modified DSLOC Only</td>
<td><a href="#">GMSdUTbl</a></td>
<td>$\exp(-3.466 \times FYIVal(@STbl+i,2022)) \times (FYTot(@GSdUTbl+i) - (FYIVal(@STbl+i,2020) - FYIVal(@STbl+i,2021))) + (FYIVal(@STbl+i,2020) - FYIVal(@STbl+i,2021))$ [DEGM8]</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Example Software Item with Unmodified DSLOC Only</td>
<td><a href="#">GMSdUTbl</a></td>
<td>$\exp(-3.466 \times FYIVal(@STbl+i,2022)) \times (FYTot(@GSdUTbl+i) - (FYIVal(@STbl+i,2020) - FYIVal(@STbl+i,2021))) + (FYIVal(@STbl+i,2020) - FYIVal(@STbl+i,2021))$ [DEGM8]</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
Table 10  Growth Adjusted New DSLOC Distributions Table

<table>
<thead>
<tr>
<th>WBS/CES Description</th>
<th>Unique ID</th>
<th>Equation / Throughput</th>
<th>WBS Indent Level</th>
<th>PE Position in Distribution</th>
<th>Distribution Form</th>
<th>Grouping</th>
<th>Group Strength</th>
<th>Random Seed</th>
<th>CDF Keyword</th>
<th>Ri (+) Row Index</th>
<th>i (+) Row Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example Software Item with New, Modified, and Unmodified DSLOC</td>
<td>GS1NTH1</td>
<td>FYIVal(@CDERTbl+FYIVal(@STbl+i,2001),2018)*FYIVal(@STbl+i,2013)*FYIVal(@CDERTbl+FYIVal(@STbl+i,2001),2017)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example Software Item with New DSLOC Only</td>
<td>GS2NTH1</td>
<td>FYIVal(@CDERTbl+FYIVal(@STbl+i,2001),2018)*FYIVal(@STbl+i,2013)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example Software Item with Modified DSLOC Only</td>
<td>GS3NTH1</td>
<td>FYIVal(@CDERTbl+FYIVal(@STbl+i,2001),2018)*FYIVal(@STbl+i,2013)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example Software Item with Unmodified DSLOC Only</td>
<td>GS4NTH1</td>
<td>FYIVal(@CDERTbl+FYIVal(@STbl+i,2001),2018)*FYIVal(@STbl+i,2013)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 11  Growth Adjusted Modified DSLOC Distributions Table

<table>
<thead>
<tr>
<th>WBS/CES Description</th>
<th>Unique ID</th>
<th>Equation / Throughput</th>
<th>WBS Indent Level</th>
<th>PE Position in Distribution</th>
<th>Distribution Form</th>
<th>Grouping</th>
<th>Group Strength</th>
<th>Random Seed</th>
<th>CDF Keyword</th>
<th>Ri (+) Row Index</th>
<th>i (+) Row Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example Software Item with New, Modified, and Unmodified DSLOC</td>
<td>GS5NTH1</td>
<td>FYIVal(@CDERTbl+FYIVal(@STbl+i,2001),2020)*FYIVal(@STbl+i,2013)*FYIVal(@CDERTbl+FYIVal(@STbl+i,2001),2017)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example Software Item with New DSLOC Only</td>
<td>GS6NTH1</td>
<td>FYIVal(@CDERTbl+FYIVal(@STbl+i,2001),2020)*FYIVal(@STbl+i,2013)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example Software Item with Modified DSLOC Only</td>
<td>GS7NTH1</td>
<td>FYIVal(@CDERTbl+FYIVal(@STbl+i,2001),2020)*FYIVal(@STbl+i,2013)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example Software Item with Unmodified DSLOC Only</td>
<td>GS8NTH1</td>
<td>FYIVal(@CDERTbl+FYIVal(@STbl+i,2001),2020)*FYIVal(@STbl+i,2013)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 12 Growth Adjusted Unmodified DSLOC Distributions Table

<table>
<thead>
<tr>
<th>WBS/CES Description</th>
<th>Unique ID</th>
<th>Equation / Throughput</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example Software Item with New, Modified, and Unmodified DSLOC Only</td>
<td>STbl 0</td>
<td>$\text{FYIVal}(@\text{CDERTbl}+\text{FYIVal}(@\text{STbl}+i,2001),2022)\cdot(\text{FYIVal}(@\text{STbl}+i,2020)-\text{FYIVal}(@\text{STbl}+i,2021))/(\text{FYIVal}(@\text{STbl}+i,2036))$</td>
</tr>
<tr>
<td>Example Software Item with New DSLOC Only</td>
<td>STbl 0</td>
<td>$\text{FYIVal}(@\text{CDERTbl}+\text{FYIVal}(@\text{STbl}+i,2001),2022)\cdot(\text{FYIVal}(@\text{STbl}+i,2020)-\text{FYIVal}(@\text{STbl}+i,2021))/(\text{FYIVal}(@\text{STbl}+i,2036))$</td>
</tr>
<tr>
<td>Example Software Item with Modified DSLOC Only</td>
<td>STbl 0</td>
<td>$\text{FYIVal}(@\text{CDERTbl}+\text{FYIVal}(@\text{STbl}+i,2001),2022)\cdot(\text{FYIVal}(@\text{STbl}+i,2020)-\text{FYIVal}(@\text{STbl}+i,2021))/(\text{FYIVal}(@\text{STbl}+i,2036))$</td>
</tr>
<tr>
<td>Example Software Item with Unmodified DSLOC Only</td>
<td>STbl 0</td>
<td>$\text{FYIVal}(@\text{CDERTbl}+\text{FYIVal}(@\text{STbl}+i,2001),2022)\cdot(\text{FYIVal}(@\text{STbl}+i,2020)-\text{FYIVal}(@\text{STbl}+i,2021))/(\text{FYIVal}(@\text{STbl}+i,2036))$</td>
</tr>
</tbody>
</table>

### Table 13 Software Item-Specific Inputs, Parameters, and Controls Table

<table>
<thead>
<tr>
<th>WBS/CES Description</th>
<th>Unique ID</th>
<th>Equation / Throughput</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example Software Item with New, Modified, and Unmodified DSLOC Only</td>
<td>STbl 0</td>
<td>$\text{FYIVal}(@\text{CDERTbl}+\text{FYIVal}(@\text{STbl}+i,2001),2022)\cdot(\text{FYIVal}(@\text{STbl}+i,2020)-\text{FYIVal}(@\text{STbl}+i,2021))/(\text{FYIVal}(@\text{STbl}+i,2036))$</td>
</tr>
<tr>
<td>Example Software Item with New DSLOC Only</td>
<td>STbl 0</td>
<td>$\text{FYIVal}(@\text{CDERTbl}+\text{FYIVal}(@\text{STbl}+i,2001),2022)\cdot(\text{FYIVal}(@\text{STbl}+i,2020)-\text{FYIVal}(@\text{STbl}+i,2021))/(\text{FYIVal}(@\text{STbl}+i,2036))$</td>
</tr>
<tr>
<td>Example Software Item with Modified DSLOC Only</td>
<td>STbl 0</td>
<td>$\text{FYIVal}(@\text{CDERTbl}+\text{FYIVal}(@\text{STbl}+i,2001),2022)\cdot(\text{FYIVal}(@\text{STbl}+i,2020)-\text{FYIVal}(@\text{STbl}+i,2021))/(\text{FYIVal}(@\text{STbl}+i,2036))$</td>
</tr>
</tbody>
</table>

### Table 14 JCDER Library Table

<table>
<thead>
<tr>
<th>WBS/CES Description</th>
<th>Unique ID</th>
<th>Equation / Throughput</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example Software Item with New, Modified, and Unmodified DSLOC Only</td>
<td>STbl 0</td>
<td>$\text{FYIVal}(@\text{CDERTbl}+\text{FYIVal}(@\text{STbl}+i,2001),2022)\cdot(\text{FYIVal}(@\text{STbl}+i,2020)-\text{FYIVal}(@\text{STbl}+i,2021))/(\text{FYIVal}(@\text{STbl}+i,2036))$</td>
</tr>
<tr>
<td>Example Software Item with New DSLOC Only</td>
<td>STbl 0</td>
<td>$\text{FYIVal}(@\text{CDERTbl}+\text{FYIVal}(@\text{STbl}+i,2001),2022)\cdot(\text{FYIVal}(@\text{STbl}+i,2020)-\text{FYIVal}(@\text{STbl}+i,2021))/(\text{FYIVal}(@\text{STbl}+i,2036))$</td>
</tr>
<tr>
<td>Example Software Item with Modified DSLOC Only</td>
<td>STbl 0</td>
<td>$\text{FYIVal}(@\text{CDERTbl}+\text{FYIVal}(@\text{STbl}+i,2001),2022)\cdot(\text{FYIVal}(@\text{STbl}+i,2020)-\text{FYIVal}(@\text{STbl}+i,2021))/(\text{FYIVal}(@\text{STbl}+i,2036))$</td>
</tr>
</tbody>
</table>

### Table 15 JCDER Library Table

<table>
<thead>
<tr>
<th>WBS/CES Description</th>
<th>Unique ID</th>
<th>Equation / Throughput</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example Software Item with New, Modified, and Unmodified DSLOC Only</td>
<td>STbl 0</td>
<td>$\text{FYIVal}(@\text{CDERTbl}+\text{FYIVal}(@\text{STbl}+i,2001),2022)\cdot(\text{FYIVal}(@\text{STbl}+i,2020)-\text{FYIVal}(@\text{STbl}+i,2021))/(\text{FYIVal}(@\text{STbl}+i,2036))$</td>
</tr>
<tr>
<td>Example Software Item with New DSLOC Only</td>
<td>STbl 0</td>
<td>$\text{FYIVal}(@\text{CDERTbl}+\text{FYIVal}(@\text{STbl}+i,2001),2022)\cdot(\text{FYIVal}(@\text{STbl}+i,2020)-\text{FYIVal}(@\text{STbl}+i,2021))/(\text{FYIVal}(@\text{STbl}+i,2036))$</td>
</tr>
<tr>
<td>Example Software Item with Modified DSLOC Only</td>
<td>STbl 0</td>
<td>$\text{FYIVal}(@\text{CDERTbl}+\text{FYIVal}(@\text{STbl}+i,2001),2022)\cdot(\text{FYIVal}(@\text{STbl}+i,2020)-\text{FYIVal}(@\text{STbl}+i,2021))/(\text{FYIVal}(@\text{STbl}+i,2036))$</td>
</tr>
</tbody>
</table>
Appendix F DEGM88SV Measures, Parameters, and Statistics

JCDER501: All Growth Eligible

<table>
<thead>
<tr>
<th>DSLOC Estimate Growth Model Version:</th>
<th>Version 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version 8 DSLOC Estimate Growth Model Regression Method:</td>
<td>ODR</td>
</tr>
</tbody>
</table>

DSLOC Estimate Growth Model Equations and Variables

**New DSLOC Growth Equation:**

\[ S[DGANew] \triangleq \exp(-(Decay \times Maturity)) \times (b[GN] \times \varepsilon[GN]) \times \frac{S[DNew]}{K[N]} \times a[GN] \times K[N] \]

**Modified DSLOC Growth Equation:**

\[ S[DGAMod] \triangleq \exp(-(Decay \times Maturity)) \times (b[GM] \times \varepsilon[GM]) \times \frac{S[DMod]}{K[M]} \times a[GM] \times K[M] \]

**Unmodified DSLOC Growth Equation:**

\[ S[DGAmUnmod] \triangleq \exp(-(Decay \times Maturity)) \times (b[GU] \times \varepsilon[GU]) \times \frac{S[DUnmod]}{K[U]} \times a[GU] \times K[U] \]

where:

<table>
<thead>
<tr>
<th>( a[GN] )</th>
<th>1.074</th>
<th>( a[GM] )</th>
<th>0.709</th>
<th>( a[GU] )</th>
<th>1.265</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Decay[GN] )</td>
<td>0.088</td>
<td>( Decay[GM] )</td>
<td>0.333</td>
<td>( Decay[GU] )</td>
<td>2.336</td>
</tr>
</tbody>
</table>

**List Statistics**

<table>
<thead>
<tr>
<th>Number of Data Points (observations):</th>
<th>([GN])</th>
<th>([GM])</th>
<th>([GU])</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geometric (log space) mean of ( b ):</td>
<td>5.187E-01</td>
<td>1.421E+01</td>
<td>5.016E-02</td>
</tr>
<tr>
<td>Arithmetic (unit space) mean of ( b ):</td>
<td>5.656E-01</td>
<td>2.296E+01</td>
<td>6.070E-02</td>
</tr>
<tr>
<td>Standard deviation of ( b ):</td>
<td>2.510E-01</td>
<td>2.959E+01</td>
<td>3.430E-02</td>
</tr>
<tr>
<td>Coefficient of Variation (CV) ( b ):</td>
<td>0.44</td>
<td>1.29</td>
<td>0.57</td>
</tr>
<tr>
<td>Geometric (log space) mean of ( \varepsilon ):</td>
<td>1.101E+00</td>
<td>1.769E+00</td>
<td>2.152E+00</td>
</tr>
<tr>
<td>Arithmetic (unit space) mean of ( \varepsilon ):</td>
<td>5.436E-01</td>
<td>2.205E+00</td>
<td>3.060E+00</td>
</tr>
<tr>
<td>Standard deviation of ( \varepsilon ):</td>
<td>0.49</td>
<td>1.25</td>
<td>1.42</td>
</tr>
<tr>
<td>Coefficient of Variation (CV) ( \varepsilon ):</td>
<td>110%</td>
<td>177%</td>
<td>215%</td>
</tr>
</tbody>
</table>

New to Modified DSLOC Correlation: 5.86E-01
New to Unmodified DSLOC Growth Correlation: -8.29E-02

<table>
<thead>
<tr>
<th>Growth Factor Estimating Relationships Behavior</th>
<th>New DSLOC Growth</th>
<th>Modified DSLOC Growth</th>
<th>Unmodified DSLOC Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implied Growth Factor at data set mean baseline DSLOC:</td>
<td>23% at 119,750 DSLOC</td>
<td>-37% at 45,778 DSLOC</td>
<td>-3% at 70,527 DSLOC</td>
</tr>
<tr>
<td>Implied Growth Factor at data set geometric mean baseline DSLOC:</td>
<td>16% at 55,026 DSLOC</td>
<td>-10% at 13,140 DSLOC</td>
<td>-10% at 52,272 DSLOC</td>
</tr>
</tbody>
</table>

Mean Magnitude of the Relative Error:
- 110% at 52,272 DSLOC
- 177% at 119,750 DSLOC
- 215% at 45,778 DSLOC

Arithmetic (unit space) mean of \( \varepsilon \):
- 1.25 at 55,026 DSLOC
- 1.42 at 119,750 DSLOC
- 0.49 at 45,778 DSLOC

Coefficient of Variation (CV) of \( \varepsilon \):
- 10% at 52,272 DSLOC
- 16% at 119,750 DSLOC
- 23% at 45,778 DSLOC

Standard deviation of \( \varepsilon \):
- 2.205E+00 at 13,140 DSLOC
- 2.296E+01 at 119,750 DSLOC
- 2.510E-01 at 52,272 DSLOC
### Variables and Notation

#### Variables Used in JCDER and DSLOC Estimate Growth Equations

- **S** ≡ Size (ESLOC) distribution
- **E** ≡ Core Effort (person-months) or (pm) distribution (data set SDLC)
- **T** ≡ Core Duration (calendar months) or (cm) distribution (data set SDLC)
- **E[A-A]** ≡ All-All Effort (person-months) or (pm) distribution (All-All SDLC)
- **T[A-A]** ≡ All-All Duration (calendar months) or (cm) distribution (All-All SDLC)

#### Decay Parameters

- **alpha[E]** ≡ Effort dimension exponent (economy/diseconomy of scale)
- **alpha[T]** ≡ Duration dimension exponent (economy/diseconomy of scale)
- **alpha[S]** ≡ Size dimension exponent (economy/diseconomy of scale)

#### Decay Factors

- **D** ≡ Difficulty (nonlinear person-month months per ESLOC) distribution -- time-sensitive inverse productivity
- **I** ≡ Intensity (nonlinear person-months per calendar month) distribution -- staffing or burn

#### Estimation Growth Factors

- **Eaf** ≡ Effort Adjustment Factor (data set SDLC to All-All SDLC)
- **Daf** ≡ Duration Adjustment Factor (data set SDLC to All-All SDLC)

#### Software Item (SI) to CSCI Normalization Factor

- **K[S]** ≡ Software Item (SI) being estimated to CSCI normalization factor for ESLOC
- **K[GN]** ≡ New DSLOC estimating relationship exponent (economy/diseconomy of scale)
- **K[GM]** ≡ Modified DSLOC estimating relationship exponent (economy/diseconomy of scale)
- **K[GU]** ≡ Unmodified DSLOC estimating relationship exponent (economy/diseconomy of scale)

#### Baseline Estimate Growth Error Factors

- **ε[GN]** ≡ Baseline New DSLOC estimate growth error factor distribution
- **ε[GM]** ≡ Baseline Modified DSLOC estimate growth error factor distribution
- **ε[GU]** ≡ Baseline Unmodified DSLOC estimate growth error factor distribution

#### Technical Baseline Estimates

- **S[DN]** ≡ Technical Baseline Estimate (TBE) of New DSLOC
- **S[DM]** ≡ Technical Baseline Estimate (TBE) of Modified DSLOC
- **S[DU]** ≡ Technical Baseline Estimate (TBE) of Unmodified DSLOC

#### Growth-Adjusted Estimates

- **S[DGA]** ≡ Growth-adjusted DSLOC estimate distributions
- **S[DGANO]** ≡ Growth-adjusted New DSLOC estimate distribution
- **S[DGAMO]** ≡ Growth-adjusted Modified DSLOC estimate distribution
- **S[DGAU]** ≡ Growth-adjusted Unmodified DSLOC estimate distribution

### Notation Conventions

- ‘≡’ is used to indicate identity; the left expression is defined by right expression
- ‘≜’ is used to indicate estimation; the right expression estimates the left expression

**Arial Bold Italic** font indicates a random variable; range and distribution of possible outcomes

**Times New Roman Italic** font indicates a specific-value (single-value) variable

**Bracks ‘[ ]’** are used to indicate the subscripted characters of a variable name
<table>
<thead>
<tr>
<th>ODR -- SVDD(M[Est,Act])</th>
<th>ODR -- SVDV(M[Est,Act])</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.517E+00 0.000E+00 0.000E+00 1.074</td>
<td>9.373E-01 1.109E-02 7.323E-02</td>
</tr>
<tr>
<td>0.000E+00 5.559E+00 0.000E+00 2.456E+00</td>
<td>0.000E+00 9.973E-01 1.109E-02</td>
</tr>
<tr>
<td>0.000E+00 0.000E+00 9.973E-01</td>
<td>1.109E-02 6.517E+00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>a(GN) (geometric mean)</th>
<th>b(GN) data set arithmetic mean</th>
<th>b(GN) data set standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.187E-01 0.000E+00 0.000E+00 5.659E-01</td>
<td>2.510E-01 0.000E+00 9.973E-01</td>
<td></td>
</tr>
<tr>
<td>0.000E+00 5.559E+00 0.000E+00</td>
<td>1.109E-02 -7.323E-02 5.559E+00</td>
<td></td>
</tr>
<tr>
<td>0.000E+00 0.000E+00 9.973E-01</td>
<td>1.109E-02 6.517E+00</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>b(GN) data set CV</th>
<th>Generalized R²</th>
<th>%SOEE Unit Space</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.44</td>
<td>0.90</td>
<td>49.94%</td>
</tr>
<tr>
<td>b(GN) data set mean</td>
<td>Growth @ min(SDNAct)</td>
<td>0.000E+00</td>
</tr>
<tr>
<td>1.101E+00</td>
<td>Growth @ max(SDNAct)</td>
<td>0.000E+00</td>
</tr>
<tr>
<td>5.438E-01</td>
<td>Relationship MMROE</td>
<td>110%</td>
</tr>
<tr>
<td>e(GN) data set CV</td>
<td>Decay</td>
<td>0.088</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Implied growth factor at data set mean size</th>
<th>23% at 119,750 DSLOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implied growth factor at data set geometric mean size</td>
<td>16% at 55,026 DSLOC</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ODR -- SVDD(M[Est,Act])</th>
<th>ODR -- SVDV(M[Est,Act])</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.338E+00 0.000E+00 0.000E+00 7.090</td>
<td>9.535E-01 -2.241E-01 2.895E-01</td>
</tr>
<tr>
<td>0.000E+00 5.184E+00 0.000E+00</td>
<td>-2.861E-01 -6.154E-02 5.184E+00</td>
</tr>
<tr>
<td>0.000E+00 0.000E+00 3.977E+00</td>
<td>-6.154E-02 9.954E-01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>a(GM) (geometric mean)</th>
<th>b(GM) data set arithmetic mean</th>
<th>b(GM) data set standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.421E+01 0.000E+00 0.000E+00 2.296E+01</td>
<td>2.959E+01 0.000E+00 9.535E-01</td>
<td></td>
</tr>
<tr>
<td>0.000E+00 5.184E+00 0.000E+00</td>
<td>-2.861E-01 -6.154E-02 5.184E+00</td>
<td></td>
</tr>
<tr>
<td>0.000E+00 0.000E+00 3.977E+00</td>
<td>-6.154E-02 9.954E-01</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>b(GM) data set CV</th>
<th>%SOEE Unit Space</th>
<th>min(SDNAct)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.29</td>
<td>64.67%</td>
<td>2.00</td>
</tr>
<tr>
<td>b(GM) data set mean</td>
<td>Growth @ min(SDNAct)</td>
<td>max(SDNAct)</td>
</tr>
<tr>
<td>1.769E+00</td>
<td>Growth @ max(SDNAct)</td>
<td>240.100</td>
</tr>
<tr>
<td>2.205E+00</td>
<td>Relationship MMROE</td>
<td>177%</td>
</tr>
<tr>
<td>e(GM) data set standard deviation</td>
<td>Decay</td>
<td>0.333</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Implied growth factor at data set mean size</th>
<th>-37% at 45,778 DSLOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implied growth factor at data set geometric mean size</td>
<td>-10% at 13,140 DSLOC</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ODR -- SVDD(M[Est,Act])</th>
<th>ODR -- SVDV(M[Est,Act])</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.305E+00 0.000E+00 0.000E+00 1.265</td>
<td>6.660E-01 -6.875E-01 2.895E-01</td>
</tr>
<tr>
<td>0.000E+00 4.892E+00 0.000E+00</td>
<td>-2.933E-01 1.152E-01 9.490E-01</td>
</tr>
<tr>
<td>0.000E+00 0.000E+00 2.268E+00</td>
<td>-6.857E-01 -7.170E-01 1.251E-01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>a(GU) (geometric mean)</th>
<th>b(GU) data set arithmetic mean</th>
<th>b(GU) data set standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.016E-02 0.000E+00 0.000E+00 3.430E-02</td>
<td>3.430E-02 0.000E+00 9.535E-01</td>
<td></td>
</tr>
<tr>
<td>0.000E+00 4.892E+00 0.000E+00</td>
<td>-2.861E-01 -6.154E-02 5.184E+00</td>
<td></td>
</tr>
<tr>
<td>0.000E+00 0.000E+00 2.268E+00</td>
<td>-6.154E-02 9.954E-01</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>b(GU) data set CV</th>
<th>%SOEE Unit Space</th>
<th>min(SDUAct)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.29</td>
<td>49.11%</td>
<td>8.000</td>
</tr>
<tr>
<td>b(GU) data set mean</td>
<td>Growth @ min(SDUAct)</td>
<td>max(SDUAct)</td>
</tr>
<tr>
<td>1.769E+00</td>
<td>Growth @ max(SDUAct)</td>
<td>213.010</td>
</tr>
<tr>
<td>2.205E+00</td>
<td>Relationship MMROE</td>
<td>215%</td>
</tr>
<tr>
<td>e(GU) data set standard deviation</td>
<td>Decay</td>
<td>2.336</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Implied growth factor at data set mean size</th>
<th>-3% at 70,527 DSLOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implied growth factor at data set geometric mean size</td>
<td>-10% at 52,722 DSLOC</td>
</tr>
</tbody>
</table>

### Correlation Matrix

<table>
<thead>
<tr>
<th></th>
<th>New DSLOC</th>
<th>Modified DSLOC</th>
<th>Unmodified DSLOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>New DSLOC</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modified DSLOC</td>
<td>5.867595E-01</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Unmodified DSLOC</td>
<td>-8.293332E-02</td>
<td>-8.104507E-02</td>
<td>1</td>
</tr>
</tbody>
</table>
Figure 14  SV New DSLOC data regressed with ODR in log space (DEGM8SV) yields a better fit than does simply using the DEGM7 relationship.

Figure 15  SV Modified DSLOC data regressed with ODR in log space (DEGM8SV) yields a better fit than does simply using the DEGM7 relationship.
**Figure 16**  SV Unmodified DSLOC data regressed with ODR in log space (DEGM8SV) yields a better fit than does simply using the DEGM7 relationship.
NOTES

1 The term custom CDF refers to a feature in Tecolote Research, Inc.’s ACE software tool that allows distributions to be specified as a discrete range-value-to-percentile mapping as opposed to a mapping described by some mathematical distribution function such as “lognormal”.

2 We use the Arial bold italic font to denote a random variable; i.e., a variable that can take on values according to some probability distribution, the Times New Roman bold italic font to denote a function, the Times New Roman bold font to denote a vector or matrix or list or array, the Times New Roman italic font to denote a simple variable, and the Times New Roman normal font to denote a number. We use the overstrike caret (^) character to indicate an object that represents an estimated value. We use the overstrike tilde (~) character to indicate an object that represents the geometric mean value (arithmetic mean value in log space) of an associated list or random variable (acts as a descriptor), or the geometric mean value of a list or random variable (acts as a function); depending on context.

3 We use the term attainment probability to describe the probability that the actual outcome will be less than or equal to a particular value in the distribution of outcomes; i.e., a particular outcome’s percentile rank.

4 These percentages are default values representing rough approximations of the percentages used by various software estimating models.

5 Note that it is possible to expand the estimate Maturity input to include specific (and possibly different) Maturity values for New, Modified, and Unmodified DSLOC. This can be done as part of the model implementation since each of New, Modified, and Unmodified DSLOC has its own unique growth equation.

6 SDLCBegin assumes zero maturity.

7 This data set, circa 2015, contains 2,861 observations (submitted DD Form 2630’s).

8 Linear algebra defines the standard basis in \( \mathbb{R}^m \) as an ordered set of \( m \) distinct standard unit vectors \( \{ \hat{e}_1, \hat{e}_2, \ldots, \hat{e}_n \} \) where each \( \hat{e}_i \) denotes a vector with the value of 1 as its \( i \)-th coordinate and the value of 0 as every other coordinate. Locating the initial points of each standard unit vector \( \hat{e}_i \) at the \( \mathbb{R}^m \) reference origin \( P_0 \) serves to locate all the positive coordinate axes in \( \mathbb{R}^m \).

9 Physical quantities that are commensurable have the same dimension and can be directly compared to each other, even if they are originally expressed in differing units of measure (such as inches and meters, or pounds and newtons). If physical quantities have different dimensions (such as length vs. mass), they cannot be expressed in terms of similar units and cannot be compared in quantity (also called incommensurable). For example, asking whether a kilogram is greater than, equal to, or less than an hour is meaningless.

[https://en.wikipedia.org/wiki/Dimensional_analysis]
In physics, mathematics, statistics, and economics, *scale invariance* is a feature of objects or laws that do not change if scales of length, energy, or other variables, are multiplied by a common factor, thus represent a universality.

10 [https://en.wikipedia.org/wiki/Total_least_squares]

11 [https://en.wikipedia.org/wiki/Distance_from_a_point_to_a_line]

12 In regression analysis, the difference between the observed value of the dependent variable and its predicted value is called the residual.


13 The convolution of probability distributions arises in probability theory and statistics as the operation in terms of probability distributions that corresponds to the addition of independent random variables and, by extension, to forming linear combinations of random variables.

[https://en.wikipedia.org/wiki/Convolution_of_probability_distributions]


15 Not to be confused with use of the overstrike tilde (~) notation to indicate that a value is a geometric mean as is done in the body of the paper.

16 We use the overstrike caret or *hat* (^) notation on a vector to indicate that it is a unit vector.

17 [https://en.wikipedia.org/wiki/Principal_component_analysis]

**REFERENCES**


and Analysis, 2005.
ABOUT THE AUTHORS

Eric Sommer has a Bachelor of Science degree in Mathematics and a Finance focused MBA. He began his career as a high school math teacher, teaching algebra to AP Calculus, before joining the cost estimating community at the Space and Missile Systems Center (SMC). He currently is an Acting Cost Chief and an Operations Research Analyst at SMC focusing on software cost estimating and office automation.

Bopha Seng is a senior analyst with Tecolote Research supporting the Cost Estimating and Research Divisions at the Space and Missile Systems Center (SMC). Her primary responsibilities include space and ground cost estimating/modelling and methodology development. Prior to Tecolote, she was a civilian at SMC where she was the MILSATCOM advanced concepts cost lead and POM lead. Bopha is a CCE/A and has a BA in Statistics from UC Berkeley.

David LaPorte is an analyst with Tecolote Research supporting the Cost Research Division at the Space and Missile Systems Center. His chief research efforts include data collections, space vehicle mass growth analysis, and software growth analysis. He has been working with Tecolote for 6 months after graduating top of his class from California State University Dominguez Hills with a degree in Finance.

Michael Ross has 40 years of experience in software engineering as a developer, manager, consultant, instructor, and award-winning international speaker. Mr. Ross is currently President and CEO of r2Estimating, LLC (developers of the r2-v2 Software Estimating Framework). Mr. Ross is a Life Member of ICEAA and regularly presents papers at its annual conferences (four of which have been recognized with Best Paper Awards). Mr. Ross has a BS in Computer Engineering from Arizona State University.