
Ray Madachy
Naval Postgraduate School

Brad Clark
USC Center for Systems and Software Engineering

USC CSSE Annual Research Review 2015
April 16, 2015
Overview

• DoD Systems Engineering Research Center (SERC) research transition e-book:


• Downloadable from
  – Book site [http://softwarecost.org](http://softwarecost.org)
  – Other sites forthcoming
Over 1,500 hits as of Apr 15
Contributors

• **Barry Boehm**, University of Southern California
• **Bradford Clark**, Software Metrics Inc.
• **Joseph Dean**, Air Force Cost Analysis Agency
• **Cheryl Jones**, US Army Armament Research Development and Engineering Center
• **Raymond Madachy**, Naval Postgraduate School
• **John McGarry**, US Army Armament Research Development and Engineering Center
• **Wilson Rosa**, Naval Center for Cost Analysis
Acknowledgments

• Numerous workshop participants during the research

• External reviewers
  – Jo Ann Lane, University of Southern California
  – Daniel Ligett, Softstar Systems
  – Daniel Nussbaum, Naval Postgraduate School
  – David Seaver, National Security Agency
  – Richard Selby, Northrop Grumman
  – Daniel Strickland, Missile Defense Agency
  – David Zubrow, Software Engineering Institute

• Cost estimation tool companies providing support
  – Galorath Systems
  – PRICE Systems
  – Quantitative Software Management
  – Softstar Systems
Table of Contents

1 Introduction
2 Software Resources Data Report
3 Metrics Definitions
4 Data Assessment
5 Cost Estimating Relationships
6 Modern Estimating Challenges
7 Estimation Process
Appendix A - Cost Model Descriptions
Appendix B - MIL-STD-881C WBS Mapping
Appendix C - Unified CodeCount
Appendix D – Acronyms
Bibliography
Index
Purpose

- Assist analysts and decision makers early software cost estimates for different types of DoD systems and operating environments

- The intent is to improve quality and consistency of early software estimating through guidance, standardization, and knowledge sharing.

- We have analyzed DoD empirical software cost data and are transitioning the results back in this open access manual.

- We describe our processes for data normalization, analysis, derivation of Cost Estimating Relationships (CERs) and productivity benchmarking.
Estimation and Metrics Processes

Defense Acquisition Management System

Material Solution Analysis  Technology Development  Engineering and Manufacturing Development  Production and Deployment  Operations and Support

SRDRs

Simple CER Estimates  Multi-Parameter CER Estimates

Program Estimation Process

Initiation and Research  Assessment  Analysis  Documentation and Presentation

Simple CERs  Multi-Parameter CERs

Software Cost Estimation Metrics Process

Collect Data  Prepare Data  Create CERs  Address Challenges

SRDRs

Metrics Definitions  Cost Model Descriptions

Evolve CERs
Ch-2. Software Resources Data Report

• Overview
• Collecting Organization
• Repository
• Reporting Frequency
• SRDR Content
• Further SRDR Resources
Dataset Summary

- Analysis based on 317 SRDRs reported from recent DoD projects during 2004-2013.
- Normalized Effort vs. Equivalent Size in SRDRs:

![Graph showing the relationship between KESLOC and Person-Months for different military branches: Navy, Army, and Air Force.]
Ch-3. Metrics Definitions

• Overview
• Product Size Measures
• SLOC Count Definitions
• ESLOC Summary
• Functional Size Measures
• Development Effort
• Development Duration
Metrics Definitions Examples

• Software Size Types

<table>
<thead>
<tr>
<th>Size Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>New</td>
<td>New software created for the first time.</td>
</tr>
<tr>
<td>Adapted</td>
<td>Pre-existing software that is used as-is (Reused) or changed (Modified).</td>
</tr>
<tr>
<td>Reused</td>
<td>Pre-existing software that is not changed with the adaption parameter settings:</td>
</tr>
<tr>
<td></td>
<td>• Design Modification % (DM) = 0%</td>
</tr>
<tr>
<td></td>
<td>• Code Modification % (CM) = 0%.</td>
</tr>
<tr>
<td>Modified</td>
<td>Pre-existing software that is modified for use by making design, code and / or test changes:</td>
</tr>
<tr>
<td></td>
<td>• Code Modification % (CM) &gt; 0%.</td>
</tr>
</tbody>
</table>

• SLOC Count Definitions

<table>
<thead>
<tr>
<th>Statement Type</th>
<th>Logical In</th>
<th>Logical Out</th>
<th>NCSS In</th>
<th>NCSS Out</th>
<th>Physical In</th>
<th>Physical Out</th>
<th>Total In</th>
<th>Total Out</th>
</tr>
</thead>
<tbody>
<tr>
<td>Executable</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Nonexecutable

<table>
<thead>
<tr>
<th>Statement Type</th>
<th>Logical In</th>
<th>Logical Out</th>
<th>NCSS In</th>
<th>NCSS Out</th>
<th>Physical In</th>
<th>Physical Out</th>
<th>Total In</th>
<th>Total Out</th>
</tr>
</thead>
<tbody>
<tr>
<td>Declarations</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Compiler directives</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Comments</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Blank lines</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
Ch-4. Data Assessment

• Gather Collected Data
• Inspect each Data Record
• Determine Data Quality Levels
• Correct Missing or Questionable Data
• Normalize Size and Effort Data
• Convert Raw SLOC into Equivalent SLOC
Data Assessment Examples

• Data Quality Rating Scale

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Value</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size:</td>
<td>1</td>
<td>if size data present</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>if no size data</td>
</tr>
<tr>
<td>Size Count Type</td>
<td>1</td>
<td>if size is Logical SLOC</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>if size is Non-Commented Source Statements</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>if size is Physical Lines (Comment and Source Statements)</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>if size is Total Lines (all lines in file: blank, comment, source)</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>if no size data</td>
</tr>
<tr>
<td>ESLOC Factors: (See Section 3.2.2.2)</td>
<td>1</td>
<td>if modification factors are provided for Autogen, Modified &amp; Reuse code counts from out-</td>
</tr>
</tbody>
</table>

• Software reuse proxy values for DM, CM and IM

<table>
<thead>
<tr>
<th>Code Type</th>
<th>#</th>
<th>DM Range</th>
<th>CM Range</th>
<th>IM Range</th>
<th>AAF</th>
</tr>
</thead>
<tbody>
<tr>
<td>New</td>
<td>1</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>1.0</td>
</tr>
<tr>
<td>Modified</td>
<td>101</td>
<td>0-100%</td>
<td>1-100%</td>
<td>3-100%</td>
<td>0.47</td>
</tr>
<tr>
<td>Reused</td>
<td>145</td>
<td>0%</td>
<td>0%</td>
<td>0-100%</td>
<td>0.03</td>
</tr>
<tr>
<td>Auto-Gen</td>
<td>6</td>
<td>0%</td>
<td>0%</td>
<td>0-100%</td>
<td>0.15</td>
</tr>
</tbody>
</table>
Ch-5. Cost Estimating Relationships

- Overview
- Data Segmentation
- Estimating Relationships
- SRDR-Derived CERs
- SRDR-Derived Benchmarks
- Limitations and Future Work
Data Analysis Objectives

• Make collected data useful to oversight and management entities
  – Provide guidance on how to condition data to address challenges
  – Segment data into different Application Types and Operating Environments
  – Analyze data for simple Cost Estimating Relationships (CER) within each domain
  – Develop rules-of-thumb for missing data

Data Records for one Domain
Operating Environment

- Represents the platform that the software operates in.
- Complexity of a software also driven by environment
- It is important to determine the appropriate environment before analyzing your software project:

<table>
<thead>
<tr>
<th>Environment</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ground Site (GS)</td>
<td>Fixed (GSF)</td>
</tr>
<tr>
<td></td>
<td>Mobile (GSM)</td>
</tr>
<tr>
<td>Ground Vehicle (GV)</td>
<td>Manned (GVM)</td>
</tr>
<tr>
<td></td>
<td>Unmanned (GVU)</td>
</tr>
<tr>
<td>Maritime Vessel (MV)</td>
<td>Manned (MVM)</td>
</tr>
<tr>
<td></td>
<td>Unmanned (MVU)</td>
</tr>
<tr>
<td>Aerial Vehicle (AV)</td>
<td>Manned (AVM)</td>
</tr>
<tr>
<td></td>
<td>Unmanned (AVU)</td>
</tr>
<tr>
<td>Space Vehicle (SV)</td>
<td>Manned (SVM)</td>
</tr>
<tr>
<td></td>
<td>Unmanned (SVU)</td>
</tr>
<tr>
<td>Ordnance Vehicle (OV)</td>
<td>Unmanned (OVU)</td>
</tr>
</tbody>
</table>
Application Type

• Complexity is influenced by its **Application Type**

• **Application Types** are groups of application domains that are environment independent, technology driven and characterized by:

  1. Required software reliability
  2. Data processing requirements
  3. Product complexity
  4. Integration complexity
  5. Real-time operating requirements
  6. Platform volatility
  7. Target system volatility
  8. Special display requirements
  9. Development re-hosting
  10. Quality assurance requirements
  11. Security requirements
  12. Assurance requirements
  13. Required testing level

• Determining the appropriate productivity type is critical as it allows you the select the most appropriate model and benchmark
Application Types

1. Sensor Control and Signal Processing (SCP)
2. Vehicle Control (VC)
3. Vehicle Payload (VP)
4. Real Time Embedded (RTE)
5. Mission Processing (MP)
6. Process Control (PC)
7. System Software (SYS)
8. Planning Software (PLN)
9. Scientific Software (SCI)
10. Training Software (TRN)
11. Telecommunications (TEL)
12. Software Tools (TOOL)
13. Test Software (TST)
14. Intelligence & Information Software (IIS)
Example CER: Real Time Embedded (1/2)

\[ PM = 13.20 \cdot KESLOC^{0.84} \]

Data Points = 57
KESLOC Min = 2
KESLOC Max = 201

\[ R^2 = 83\% \]
SE = 211 PM
PRED = 54%
Example CER: Real Time Embedded (2/2)

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Value</th>
<th>T-Stat</th>
<th>P-Value</th>
<th>LCI</th>
<th>UCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>13.16</td>
<td>14.7</td>
<td>1.00E-20</td>
<td>9.26</td>
<td>18.69</td>
</tr>
<tr>
<td>B</td>
<td>0.84</td>
<td>16.2</td>
<td>1.48E-22</td>
<td>0.73</td>
<td>0.94</td>
</tr>
</tbody>
</table>

\[ PM = 13.20 \cdot KESLOC^{0.84} \]

\[ \pm \text{Standard Error} \]
# Operating Environment CERs

<table>
<thead>
<tr>
<th>Op Env</th>
<th>Effort Equation</th>
<th>N</th>
<th>$R^2$</th>
<th>SE</th>
<th>PRED</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>AV</td>
<td>$PM = 9.41 \cdot KESLOC^{0.93}$</td>
<td>87</td>
<td>77%</td>
<td>516</td>
<td>39</td>
<td>0.7</td>
<td>329.9</td>
</tr>
<tr>
<td>GS</td>
<td>$PM = 12.6 \cdot KESLOC^{0.79}$</td>
<td>134</td>
<td>74%</td>
<td>478</td>
<td>56</td>
<td>0.8</td>
<td>842.1</td>
</tr>
<tr>
<td>GV</td>
<td>$PM = 15.1 \cdot KESLOC^{0.82}$</td>
<td>26</td>
<td>74%</td>
<td>178</td>
<td>54</td>
<td>7.2</td>
<td>283.1</td>
</tr>
<tr>
<td>MV</td>
<td>$PM = 5.44 \cdot KESLOC^{1.12}$</td>
<td>42</td>
<td>87%</td>
<td>206</td>
<td>33</td>
<td>0.4</td>
<td>123.8</td>
</tr>
<tr>
<td>OV</td>
<td>$PM = 27.45 \cdot KESLOC^{0.71}$</td>
<td>25</td>
<td>79%</td>
<td>11</td>
<td>48</td>
<td>0.9</td>
<td>221.1</td>
</tr>
</tbody>
</table>
## Application Type CERs

<table>
<thead>
<tr>
<th>App Type</th>
<th>Effort Equation</th>
<th>N</th>
<th>$R^2$</th>
<th>SE</th>
<th>PRED</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAS</td>
<td>$PM = 2.64 \cdot KESLOC^{1.02}$</td>
<td>16</td>
<td>97%</td>
<td>135</td>
<td>88</td>
<td>2.4</td>
<td>417.1</td>
</tr>
<tr>
<td>COM</td>
<td>$PM = 7.30 \cdot KESLOC^{0.91}$</td>
<td>47</td>
<td>88%</td>
<td>253</td>
<td>60</td>
<td>0.4</td>
<td>532.4</td>
</tr>
<tr>
<td>C&amp;C</td>
<td>$PM = 6.60 \cdot KESLOC^{1.05}$</td>
<td>33</td>
<td>88%</td>
<td>449</td>
<td>52</td>
<td>0.8</td>
<td>229.0</td>
</tr>
<tr>
<td>MP</td>
<td>$PM = 6.14 \cdot KESLOC^{0.86}$</td>
<td>20</td>
<td>77%</td>
<td>220</td>
<td>55</td>
<td>9.6</td>
<td>570.0</td>
</tr>
<tr>
<td>RTE</td>
<td>$PM = 13.20 \cdot KESLOC^{0.84}$</td>
<td>57</td>
<td>83%</td>
<td>211</td>
<td>54</td>
<td>1.8</td>
<td>200.6</td>
</tr>
<tr>
<td>SCP</td>
<td>$PM = 26.5 \cdot KESLOC^{0.87}$</td>
<td>36</td>
<td>91%</td>
<td>346</td>
<td>50</td>
<td>0.8</td>
<td>192.9</td>
</tr>
<tr>
<td>S&amp;S</td>
<td>$PM = 7.43 \cdot KESLOC^{0.91}$</td>
<td>17</td>
<td>85%</td>
<td>176</td>
<td>53</td>
<td>4.4</td>
<td>225.8</td>
</tr>
<tr>
<td>SYS</td>
<td>$PM = 5.06 \cdot KESLOC^{0.98}$</td>
<td>27</td>
<td>93%</td>
<td>585</td>
<td>48</td>
<td>0.8</td>
<td>842.1</td>
</tr>
<tr>
<td>TMDE</td>
<td>$PM = 7.42 \cdot KESLOC^{1.00}$</td>
<td>11</td>
<td>92%</td>
<td>454</td>
<td>27</td>
<td>0.5</td>
<td>313.0</td>
</tr>
<tr>
<td>VC</td>
<td>$PM = 9.05 \cdot KESLOC^{1.02}$</td>
<td>27</td>
<td>92%</td>
<td>303</td>
<td>48</td>
<td>0.7</td>
<td>329.9</td>
</tr>
<tr>
<td>VP</td>
<td>$PM = 22.27 \cdot KESLOC^{0.81}$</td>
<td>18</td>
<td>89%</td>
<td>111</td>
<td>56</td>
<td>1.1</td>
<td>221.1</td>
</tr>
</tbody>
</table>
Application Type Productivity Benchmarks
Application Type and Operating Environment Productivity Benchmarks
Ch-7. Estimation Process (Based on GAO)

- Overview
- Estimation Purpose
- Program Definition
- Estimation Scope
- Data Collection and Normalization
- Estimate Creation
- Sensitivity Analysis
- Risk and Uncertainty Analysis
- Estimate Documentation and Packaging
Risk and Uncertainty Analysis

Estimation Process

FIGURE 7.7: RTE Effort Cumulative Distribution Function
Future Work

• More SRDR data is collected each year. The SRDR data requirements are periodically revised to collect more and higher quality data.
  – The intent is to keep this manual relevant with future editions incorporating new information from contributors.

• Much future work is identified in the current manual.

• This manual is hosted and maintained at http://softwarecost.org
  – Readers will find errata, updates to its content, and a place to submit suggestions and comments.
Future Work

• The intent is to keep this manual relevant for assisting analysts and decision makers with early software cost estimates and commitments

• Future editions will incorporate new information from contributors (next slides)

• Much future work is identified in the current manual.

• This manual is hosted and maintained at http://softwarecost.org
  – Readers will find errata, updates to its content, and a place to submit suggestions and comments.
Effort Growth Contract Award to End

Boxplot of Effort Growth (Contract Award to End)

- ERP: 22.5%
- AIS: 15%
- Defense: 37.5%

Project Type
Effort Model Based on Requirements

Equation:

\[ PM = aREQ^{0.9679} \]

Where:

- \( PM \) = Actual effort (in Person Months)
- \( aREQ \) = Actual total requirements

<table>
<thead>
<tr>
<th>Model Form</th>
<th>N</th>
<th>( R^2 )</th>
<th>CV</th>
<th>Mean</th>
<th>F-stat</th>
<th>REQ Min</th>
<th>REQ Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>( PM = aREQ^{0.9679} )</td>
<td>40</td>
<td>97</td>
<td>63</td>
<td>1718</td>
<td>1659</td>
<td>35</td>
<td>12716</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coeff</th>
<th>T stat</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>aREQ</td>
<td>0.9679</td>
<td>40.7</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Actual vs. Predicted (Unit Space)
Schedule Model Based on Actual Effort

Equation:

\[ TDEV = aPM^{0.5051} \]

Where:

- \( TDEV \) = Actual Duration in Months
- \( aPM \) = Actual Effort (in Person Months)

<table>
<thead>
<tr>
<th>Model Form</th>
<th>N</th>
<th>( R^2 )</th>
<th>CV</th>
<th>Mean</th>
<th>F-stat</th>
<th>PM Min</th>
<th>PM Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>( TDEV = aPM^{0.5051} )</td>
<td>40</td>
<td>95</td>
<td>48</td>
<td>38</td>
<td>887</td>
<td>27</td>
<td>14819</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coeff</th>
<th>T stat</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( aPM )</td>
<td>0.529</td>
<td>26.14</td>
<td>0.0000</td>
</tr>
</tbody>
</table>