Agile Software Development
Cost Modeling for the US DoD

Wilson Rosa, Naval Center for Cost Analysis
Ray Madachy, Naval Postgraduate School
Bradford Clark, Software Metrics, Inc.
Barry Boehm, University of Southern California

SEI Software and Cyber Solutions Symposium
March 27, 2018
A Short History of Software Estimation Accuracy

IDPD: Incremental Development Productivity Decline
MBSSE: Model-Based Systems and Sw Engr.
COTS: Commercial Off-the-Shelf
SoS: Systems of Systems

Relative Productivity
Estimation Error

Time, Domain Understanding

Unprecedented Precedented Component-based COTS Agile SoS. Apps, Widgets, IDPD, Clouds, Security, MBSSE
Problem Statement

• In DoD, Popular Size Measures are often not available for Agile Effort Estimation at early phase
  – Function Points (FP)
  – COSMIC FP
  – Story Points
  – Source Lines of Code

• No Publicized/Empirical Agile Effort Estimation Models
• Publish Agile Effort Estimation Models for
  – Crosschecking Contractor Cost Proposals
  – Validating Independent Government Cost Estimates
• Examine the validity of using Initial Software Requirements as proxy size measure
• Develop useful cost models using early phase information
• Model calibration comparison:

<table>
<thead>
<tr>
<th>Model Type</th>
<th>Size</th>
<th>Cost Factors</th>
<th>Effort</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Initial Estimate</td>
<td>Final Actual</td>
<td>Initial Estimates</td>
</tr>
<tr>
<td>Early Phase</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Traditional</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>
Outline

• Experimental Design
• Dataset Demographics
• Productivity Benchmarks
• Agile Effort Estimation Models
• Conclusion
Experimental Design
Primary Data Collection Form

### 2011 Software Resource Data Report (SRDR) (DD Form 2630)

#### SRDR Final Developer Report

**MAJOR PROGRAM**
- **NAME:**
- **PHASE/MILESTONE:**

**REPORTING ORGANIZATION TYPE**
- **NAME/ADDRESS:**
- **DIVISION:**

**PRIME/ASSOCIATE CONTRACTOR**
- **DIRECT-REPORTING SUBCONTRACTOR**
- **GOVERNMENT**

**APPROVED PLAN NUMBER**
- **CUSTOMER**
- **CONTRACT TYPE**
- **WBS ELEMENT CODE**
- **WBS REPORTING ELEMENT**
- **TYPE ACTION**
- **NAME:**
- **SOLICITATION NO.:**
- **TASK ORDER/DELIVERY ORDER NO.:**
- **LATEST MODIFICATION:**

**PERIOD OF PERFORMANCE**
- **START DATE** (YYYYMMDD): 3/27/2018
- **END DATE** (YYYYMMDD):

**APPROPRIATION**
- **SUBMISSION NUMBER**
- **REPORT AS OF (YYYYMMDD):**
- **DATE PREPARED (YYYYMMDD):**

**NAME (Last, First, Middle Initial)**
- **Department**
- **Telephone (Include Area Code):**
- **EMAIL ADDRESS:**

**DEVELOPMENT ORGANIZATION**
- **SOFTWARE PROCESS MATURITY**
- **LEAD EVALUATOR**
- **CERTIFICATION DATE**
- **EVALUATOR AFFILIATION**

**PRECEDENTS (List up to five similar systems by the same organization or team):**

**SRDR DATA DICTIONARY FILENAME**

**COMMENTS**

---

#### SRDR Initial Developer Report

**MAJOR PROGRAM**
- **NAME:**
- **PHASE/MILESTONE:**

**REPORTING ORGANIZATION TYPE**
- **NAME/ADDRESS:**
- **DIVISION:**

**PRIME/ASSOCIATE CONTRACTOR**
- **DIRECT-REPORTING SUBCONTRACTOR**
- **GOVERNMENT**

**APPROVED PLAN NUMBER**
- **CUSTOMER**
- **CONTRACT TYPE**
- **WBS ELEMENT CODE**
- **WBS REPORTING ELEMENT**
- **TYPE ACTION**
- **NAME:**
- **SOLICITATION NO.:**
- **TASK ORDER/DELIVERY ORDER NO.:**
- **LATEST MODIFICATION:**

**PERIOD OF PERFORMANCE**
- **START DATE** (YYYYMMDD):
- **END DATE** (YYYYMMDD): 3/27/2018

**APPROPRIATION**
- **SUBMISSION NUMBER**
- **REPORT AS OF (YYYYMMDD):**
- **DATE PREPARED (YYYYMMDD):**

**NAME (Last, First, Middle Initial)**
- **Department**
- **Telephone (Include Area Code):**
- **EMAIL ADDRESS:**

**DEVELOPMENT ORGANIZATION**
- **SOFTWARE PROCESS MATURITY**
- **LEAD EVALUATOR**
- **CERTIFICATION DATE**
- **EVALUATOR AFFILIATION**

**PRECEDENTS (List up to five similar systems by the same organization or team):**

**SRDR DATA DICTIONARY FILENAME**

**COMMENTS**

---

### Estimated Development Effort

**Actual Development Effort**

**Actual Development Process**

---

### Estimated Functional Requirements

- **Estimated Functional Requirements**
- **Estimated External Interfaces**
- **Estimated Peak Staff**
- **Application Domain**
Empirical data from 20 recent US DoD Agile programs:

12 Paired SRDRs from the Cost Assessment Data Enterprise (CADE)

Each paired SRDR includes:

- SRDR Initial Developer Report (Estimates)
- SRDR Final Developer Report (Actuals)

4 additional SRDRs from CADE (SRDR Final only)

4 Agile projects from proprietary source

20 Agile projects analyzed in this study
Data Normalization and Analysis Workflow

- Dataset normalized to “account for sizing units, application complexity, and content so they are consistent for comparisons” (source: GAO)
Counting Software Requirements

Initial Functional Requirements* + Initial External Interfaces* = Initial Software Requirements

“shall” statements contained in the baseline Software Requirements Specification (SRS)

“shall” statements contained in the baseline Interface Requirements Specifications (IRS)

SRDR Initial Report + SRDR Initial Report

*Typically available before contract award
*Definitions align with IEEE std. 830-1998
1) Dataset initially mapped into 17 Application Domains*

2) Then into 4 complexity groups called Super Domains

<table>
<thead>
<tr>
<th>Application Domain</th>
<th>Super Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Software Tools</td>
<td>Mission Support (SUPP)</td>
</tr>
<tr>
<td>Training</td>
<td>Automated Information System (AIS)</td>
</tr>
<tr>
<td>Enterprise Information System</td>
<td>Engineering (ENG)</td>
</tr>
<tr>
<td>Enterprise Services</td>
<td>Real Time (RTE)</td>
</tr>
<tr>
<td>Custom AIS Software</td>
<td></td>
</tr>
<tr>
<td>Mission Planning</td>
<td></td>
</tr>
<tr>
<td>Test, Measurement, and Diagnostic Equipment</td>
<td></td>
</tr>
<tr>
<td>Scientific &amp; Simulation</td>
<td></td>
</tr>
<tr>
<td>Process Control</td>
<td></td>
</tr>
<tr>
<td>System Software</td>
<td></td>
</tr>
<tr>
<td>Command &amp; Control, Communications</td>
<td></td>
</tr>
<tr>
<td>Real Time Embedded</td>
<td></td>
</tr>
<tr>
<td>Vehicle Control/Payload</td>
<td></td>
</tr>
<tr>
<td>Signal Processing, Microcode &amp; Firmware</td>
<td></td>
</tr>
</tbody>
</table>

*New DOD policy ([http://cade.osd.mil/policy/srdr](http://cade.osd.mil/policy/srdr)) requires that Application Domains are identified for reported software activities.
# Grouping Dataset by Super Domain

## Super Domains

<table>
<thead>
<tr>
<th>Super Domains</th>
<th>Support</th>
<th>AIS</th>
<th>Engineering</th>
<th>Real Time</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aircraft</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Business</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>C4I</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>Missile</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

|               | 3       | 4   | 7           | 6         | 20    |

Top 2 Operating Environments: C4I and Aircraft
Variable Selection

1) Pairwise Correlation to select Independent Variables

2) Stepwise Analysis to select Categorical Variables
## § Model Selection Based on P-Value, lowest MMRE and CV

<table>
<thead>
<tr>
<th>Measure</th>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient of Variation</td>
<td>CV</td>
<td>Percentage expression of the standard error compared to the mean of dependent variable. A relative measure allowing direct comparison among models.</td>
</tr>
<tr>
<td>P-value</td>
<td>α</td>
<td>Level of statistical significance established through the coefficient alpha (p ≤ α).</td>
</tr>
<tr>
<td>Variance Inflation Factor</td>
<td>VIF</td>
<td>Indicates whether multi-collinearity (correlation among predictors) is present in multiple regression analysis.</td>
</tr>
<tr>
<td>Coefficient of Determination</td>
<td>R²</td>
<td>The Coefficient of Determination shows how much variation in dependent variable is explained by the regression equation.</td>
</tr>
<tr>
<td>Mean Magnitude of Relative Error</td>
<td>MMRE</td>
<td>Low MMRE is an indication of high accuracy. MMRE is defined as the sample mean (M) of the magnitude relative error (MME). MME is the absolute value of the difference between Actual and Estimated effort divided by the Actual effort, (A – E) / A</td>
</tr>
</tbody>
</table>
Dataset Demographics
Dataset by Delivery Year

# of completed Agile Projects (reported in CADE) have increased since 2014
SRDR submissions provided limited information about Agile Framework. Future SRDR submissions will require developers to describe their Agile process.
Dataset by Software Size* Range

Average software size is 704 Software Requirements

*Software Size refers to the Initial Software Requirements
Dataset by Expended Effort (in Person-Months)

Average expended effort is 409 Person-Months

*Actual Effort Hours converted into Person Months using 152 hours/month
Productivity Benchmarks
Grouping by Software Domain shows significant effect on **Agile Software Productivity**
# Productivity Comparison

## Agile vs Non-Agile

**Average Productivity***

<table>
<thead>
<tr>
<th>Size Range</th>
<th>Agile</th>
<th>Non-Agile</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-100</td>
<td>0.37</td>
<td>0.33</td>
</tr>
<tr>
<td>101-500</td>
<td>0.96</td>
<td>0.80</td>
</tr>
<tr>
<td>501-5000</td>
<td>1.97</td>
<td>1.16</td>
</tr>
<tr>
<td><strong>Composite Average</strong></td>
<td><strong>0.8</strong></td>
<td><strong>0.66</strong></td>
</tr>
</tbody>
</table>

*Initial Software Requirements per Person-Months*

When grouped by Size, Agile Software Projects appear to be more productive
Agile Effort Estimation Models
## Agile Effort Model Variables

<table>
<thead>
<tr>
<th>Name</th>
<th>Acronym</th>
<th>Type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final Effort</td>
<td>EFFORT</td>
<td>Dependent</td>
<td>Actual software engineering effort (in Person-Months) at contract completion</td>
</tr>
<tr>
<td>Initial Software Requirements</td>
<td>REQ</td>
<td>Independent</td>
<td>Sum of Initial Functional Requirements and Initial External Interface Requirements collected at contract award. Counting convention based on “shall statements”</td>
</tr>
<tr>
<td>Initial Peak Staff</td>
<td>STAFF</td>
<td>Independent</td>
<td>Estimated peak team size at contract award, measured in full-time equivalent staff</td>
</tr>
<tr>
<td>Super Domain</td>
<td>SD</td>
<td>Categorical</td>
<td>Software primary application. Four Types: Mission Support, Automated Information System (AIS), Engineering, or Real Time</td>
</tr>
</tbody>
</table>
### Agile Effort Estimation Model
(Single Variable)

**Effort** = Final Effort (in Person Months) at contract completion

**REQ** = Initial Software Requirements at contract start

**Coefficient Statistics:**

<table>
<thead>
<tr>
<th>Variable</th>
<th>P-value</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>REQ</td>
<td>0.0002</td>
<td></td>
</tr>
<tr>
<td>STAFF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**Model 1:**

Effort = 14.5 \timesREQ^{0.5009}

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>Equation Form</td>
<td>N</td>
<td>R²</td>
<td>CV</td>
<td>Mean</td>
<td>MMRE</td>
</tr>
<tr>
<td>1</td>
<td>Effort = 14.5 \timesREQ^{0.5009}</td>
<td>20</td>
<td>53</td>
<td>48</td>
<td>409</td>
<td>64</td>
</tr>
</tbody>
</table>

**Agile Estimation Model not accurate when simply using REQ as input**

---

**Actual vs. Predicted (Unit Space)**

![Graph showing actual vs. predicted data points and trend line](image-url)
Agile Effort Estimation Model
(Two Variables)

<table>
<thead>
<tr>
<th>Model</th>
<th>Equation Form</th>
<th>N</th>
<th>$R^2%$</th>
<th>CV%</th>
<th>Mean</th>
<th>MMRE %</th>
<th>REQ Min</th>
<th>REQ Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>$\text{Effort} = 6.8 \times \text{REQ}^{0.4071} \times \text{STAFF}^{0.4404}$</td>
<td>20</td>
<td>60</td>
<td>36</td>
<td>409</td>
<td>52</td>
<td>10</td>
<td>4,867</td>
</tr>
</tbody>
</table>

Effort  =  Final Effort (in Person Months) at contract completion

REQ     =  Initial Software Requirements at contract start

STAFF   =  Initial (or Estimated) Peak Staff at contract start

Coefficient Statistics:

<table>
<thead>
<tr>
<th>Variable</th>
<th>P-value</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>REQ</td>
<td>0.0015</td>
<td>1.22</td>
</tr>
<tr>
<td>STAFF</td>
<td>0.0559</td>
<td>1.22</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Agile Estimation Model improves when Peak Staff is treated with REQ
Agile Effort Estimation Model
(Three Variables)

<table>
<thead>
<tr>
<th>Mod</th>
<th>Equation Form</th>
<th>N</th>
<th>R² %</th>
<th>CV %</th>
<th>Mean</th>
<th>MMRE %</th>
<th>REQ Min</th>
<th>REQ Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Effort = 1.3 × REQ^{0.5126} × STAFF^{0.4782} × SD^{1.001}</td>
<td>20</td>
<td>81</td>
<td>22</td>
<td>409</td>
<td>32</td>
<td>10</td>
<td>4,867</td>
</tr>
</tbody>
</table>

Effort = Final Effort (in Person Months) at contract completion

REQ = Initial Software Requirements at contract start

STAFF = Initial (or Estimated) Peak Staff at contract start

SD = 1 for Mission Support Super Domain (SD)
2 for Automated Information System SD
3 for Engineering SD
4 for Real Time SD

Coefficient Statistics:

<table>
<thead>
<tr>
<th>Variable</th>
<th>P-value</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>REQ</td>
<td>0.0000</td>
<td>1.45</td>
</tr>
<tr>
<td>STAFF</td>
<td>0.0045</td>
<td>1.37</td>
</tr>
<tr>
<td>SD</td>
<td>0.0003</td>
<td>1.07</td>
</tr>
</tbody>
</table>

Agile Estimation Model more accurate when all 3 variables are added
Conclusion
Primary Findings

- Initial Software Requirements* is a valid size proxy for Software Effort Estimation Models
- Models’ accuracy improves when Peak Staff and Super Domain, are treated along with Initial Software Requirements*

<table>
<thead>
<tr>
<th>Model</th>
<th>Equation Form</th>
<th>N</th>
<th>R²%</th>
<th>CV%</th>
<th>MMRE%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Effort = 14.5 \times \text{REQ}^{0.5009}</td>
<td>20</td>
<td>53</td>
<td>48</td>
<td>64</td>
</tr>
<tr>
<td>2</td>
<td>Effort = 6.8 \times \text{REQ}^{0.4071} \times \text{STAFF}^{0.4404}</td>
<td>20</td>
<td>60</td>
<td>36</td>
<td>52</td>
</tr>
<tr>
<td>3</td>
<td>Effort = 1.3 \times \text{REQ}^{0.5126} \times \text{STAFF}^{0.4782} \times \text{SD}^{1.001}</td>
<td>20</td>
<td>81</td>
<td>22</td>
<td>32</td>
</tr>
</tbody>
</table>

*Initial Software Requirements = Initial Functional Requirements + Initial External Interfaces
The Cone of Uncertainty

3/27/2018
Model Limitations and Usefulness

- Since data was analyzed at the CSCI level, effort models may not be appropriate for projects reported at the Roll-Up Level.
- Do not use Effort Estimation Models if your input parameters are outside of the model’s dataset range.

Proposed Effort Models may be used to either crosscheck or validate contract proposals as input parameters used in the study are typically available during proposal evaluation phase.

- Applicable for both, Defense and Business Systems
- Applicable for Agile Software Projects