USING COST MODELS IN A COMMERCIAL ENVIRONMENT

ELEVENTH INTERNATIONAL FORUM ON COCOMO AND SOFTWARE COST MODELING
OCTOBER 9-11, 1996

STEVEN R. SHYMAN
BUSINESS MANAGEMENT
BOEING COMMERCIAL AIRPLANE GROUP
(206) 266-3244 EMAIL: steven.r.shyman@boeing.com
ACQUIRING SOFTWARE IN A COMMERCIAL ENVIRONMENT

- WE DO NOT PAY SUPPLIERS DIRECTLY FOR NON-RECURRING EFFORT NOR ARE WE PAID FOR NON-RECURRING EFFORT

- WE DO NOT TRACK SUPPLIER NON-RECURRING EFFORT NOR PERFORM EARNED VALUE ANALYSIS

- WE ARE NOT PAID UNTIL DELIVERY AND INCUR PENALTIES FOR LATE DELIVERY AND DEFERRED FUNCTIONALITY

- SOFTWARE ON THE AIRPLANE INCLUDES A WIDE VARIETY OF SUPPLIERS AND FUNCTIONS
CONCLUSION •

USING MODELS IN A COMMERCIAL ENVIRONMENT •

RESULTS •

OBJECTIVE/APPROACH •

OUTLINE
DATA COLLECTION AND ANALYSIS

OBJECTIVE AND APPROACH

■ OBJECTIVE
- Determine if commercially available cost models can be used to estimate BCAG software development costs
- Identify major cost drivers
- Develop estimating capability

■ APPROACH
- Survey BCAG engineers to collect data on 777-200 LRUs
- Data includes environmental factors to run several models
- Compare model estimates to estimated actuals
DATA COLLECTION AND ANALYSIS
RESULTS

- ENVIRONMENT
  - DATA FROM 62 OF APPROXIMATELY 80 LRUs
  - REPRESENTS EIGHT MAJOR FUNCTIONS
  - REPRESENTS OVER 20 SUPPLIERS
  - TOTALS OVER 2 MILLION LINES OF CODE
  - EFFORT CONCENTRATED IN FEW LARGE SYSTEMS

- COMMERCIAL MODEL EVALUATION
  - COMMERCIAL MODEL ESTIMATES DEVIATE FROM REPORTED ACTUALS RANDOMLY
  - CALIBRATION OF MODELS USING DATA SET AS A WHOLE NOT POSSIBLE
  - SOME SEGREGATED DATA SETS SHOW PROMISE
  - NO ALTERNATE MODEL FORMS SHOW PROMISE

- PERCEIVED COST DRIVERS
  - REQUIREMENTS VOLATILITY
  - TIMING CONSTRAINTS
  - COMPLEXITY
  - REAL TIME CODE
LRU DISTRIBUTION BY SIZE

Size Groupings (SLOCs)

- % of LRUs
- % of SLOCs
- % of Total LM
Parametric Model Development

Typical Software Effort Estimating Relationship:

Effort = \( a \times \text{Size}^{b} \times \text{EAF} \)

Where:
- Effort is measured in Labor Months (LM)
- Size is measured as equivalent source lines of code (EQSLOC)
- \( a \) and \( b \) are parameters generated through regression
- EAF is Environmental Adjustment Factor (product of environment factors including measures of software, personnel characteristics, development and target environment)

To re-estimate \( a \) and \( b \) use either:

\[
\text{Effort} = a \times \text{Size}^{b} \text{ or } \text{Effort}/\text{EAF} = a \times \text{Size}^{b}
\]

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>( a )</th>
<th>( b )</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplier LM</td>
<td>-1.4</td>
<td>0.77</td>
<td>0.55</td>
</tr>
<tr>
<td>Supplier LM/EAF</td>
<td>0.75</td>
<td>0.51</td>
<td>0.36</td>
</tr>
<tr>
<td>Total LM</td>
<td>-1.4</td>
<td>0.79</td>
<td>0.56</td>
</tr>
<tr>
<td>Total LM/EAF</td>
<td>0.74</td>
<td>0.53</td>
<td>0.39</td>
</tr>
</tbody>
</table>
**ALTERNATE MODELS**

Typical Software Effort Estimating Relationship:

1. \( \text{Effort} = a(\text{Size})^b \times \text{EAF} \)

Alternate Effort Estimating Relationships:

2. \( \text{Effort} = a(\text{Size})^b \times \text{EAF} \times \text{EAFMOD} \)
3. \( \text{Effort} = a(\text{Size})^b \times \text{EAF} \times \text{LRUHINT}^c \)
4. \( \text{Effort} = a(\text{Size})^b \times \text{EAF} \times \text{Ada}^d \)
5. \( \text{Effort} = a(\text{Size})^b \times \text{EAF} \times \text{LRUHINT}^c \times \text{Ada}^d \)

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total LM</td>
<td>-1.4</td>
<td>0.79</td>
<td></td>
<td></td>
<td>0.56</td>
</tr>
<tr>
<td>Total LM/EAF</td>
<td>0.74</td>
<td>0.53</td>
<td></td>
<td></td>
<td>0.39</td>
</tr>
<tr>
<td>Total LM/EAFMOD</td>
<td>0.76</td>
<td>0.68</td>
<td></td>
<td></td>
<td>0.53</td>
</tr>
<tr>
<td>Total LM/EAF</td>
<td>3.19</td>
<td>0.44</td>
<td>0.03</td>
<td></td>
<td>0.41</td>
</tr>
<tr>
<td>Total LM/EAFMOD</td>
<td>0.99</td>
<td>0.63</td>
<td>0.02</td>
<td></td>
<td>0.52</td>
</tr>
<tr>
<td>Total LM/EAF</td>
<td>2.28</td>
<td>0.49</td>
<td>0.15</td>
<td></td>
<td>0.39</td>
</tr>
<tr>
<td>Total LM/EAFMOD</td>
<td>0.91</td>
<td>0.65</td>
<td>0.17</td>
<td></td>
<td>0.53</td>
</tr>
<tr>
<td>Total LM/EAF</td>
<td>4.78</td>
<td>0.38</td>
<td>0.04</td>
<td>0.22</td>
<td>0.42</td>
</tr>
<tr>
<td>Total LM/EAFMOD</td>
<td>1.34</td>
<td>0.57</td>
<td>0.03</td>
<td>0.21</td>
<td>0.54</td>
</tr>
</tbody>
</table>
**REGRESSION RESULTS FOR SEGREGATED DATA SETS**

The two equations are:

1. Total LM = \( a(Eff \ SLOC)^b \)
2. Total LM = \( a(Eff \ SLOC)^b \) *EAF

<table>
<thead>
<tr>
<th>Equation</th>
<th>a</th>
<th>b</th>
<th>R²</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.41</td>
<td>1.13</td>
<td>.67</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>.52</td>
<td>.79</td>
<td>.52</td>
<td>15</td>
</tr>
<tr>
<td>Set 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.65</td>
<td>1.05</td>
<td>.80</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>.93</td>
<td>.87</td>
<td>.94</td>
<td>11</td>
</tr>
<tr>
<td>Set 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.45</td>
<td>.88</td>
<td>.56</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>8.29</td>
<td>.44</td>
<td>.27</td>
<td>10</td>
</tr>
<tr>
<td>Set 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.19</td>
<td>.88</td>
<td>.72</td>
<td>19</td>
</tr>
<tr>
<td>2</td>
<td>.11</td>
<td>1.06</td>
<td>.84</td>
<td>19</td>
</tr>
<tr>
<td>Set 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.59</td>
<td>.48</td>
<td>.62</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>10.94</td>
<td>.25</td>
<td>.05</td>
<td>7</td>
</tr>
<tr>
<td>Set 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2.94</td>
<td>.48</td>
<td>.29</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>3.24</td>
<td>.45</td>
<td>.32</td>
<td>7</td>
</tr>
<tr>
<td>Set 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.63</td>
<td>.75</td>
<td>.65</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>2.06</td>
<td>.54</td>
<td>.49</td>
<td>7</td>
</tr>
<tr>
<td>Set 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>60.75</td>
<td>.16</td>
<td>.02</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>77.29</td>
<td>.11</td>
<td>.017</td>
<td>10</td>
</tr>
</tbody>
</table>
USING MODELS IN A COMMERCIAL ENVIRONMENT

- Improve ability to assess supplier environments in a consistent manner relative to the models
- Develop expertise in the models
- Develop effective tools for communicating between estimator and project engineer
- Establish baselines for understanding relative accuracy and use of models
DEVELOPED INTERNALLY AT BOEING
- USED WITH ANY MODEL
- IDENTIFIES INPUTS AND THEIR POTENTIAL RANGE OF IMPACT
- COMMUNICATION TOOL BETWEEN ESTIMATOR AND ENGINEER

DEVELOP HISTORICAL BASELINES
- AID IN ADJUSTING MODEL INPUTS FOR ACCURACY

DERIVE ANALOGY
- COMPARE AND CONTRAST INPUTS
- WHAT ARE THE RELATIVE IMPACTS?

COST DRIVER ANALYSIS
- IDENTIFY MAJOR FACTORS
- SUPPORT MITIGATING/MANAGEMENT ACTIONS
ESTIMATING FACTORS FOR EFFORT – CI 1 VS CI 2

ESTIMATING DATA
- Estimated Effort
- Requirements volatility
- Application complexity
- Degree of real time
- Interface Level
- Reliability design level
- Security level
- Process modernization
- Process volatility
- Requirements Formality
- Test rigor
- Host complexity
- Degree of automation
- Resource accessibility
- Resource fragmentation
- Resource responsiveness
- Degree of rehosting
- Target complexity
- Target memory constraints
- Target timing constraints
- Team Capability
- Development Experience
- Application experience

Percent Nominal Effort

0.5 0.75 1 1.25 1.5 1.75 2
ESTIMATING FACTORS FOR EFFORT -- CI 1

Percent Nominal Effort

- Default Value
- CI 1
- Base

Estimating Factors:
- Estimated Effort
- Requirements volatility
- Application complexity
- Degree of real-time
- Interface Level
- Reusability design level
- Security level
- Process modernization
- Process variability
- Requirements Formality
- Test rigor
- Host complexity
- Degree of automation
- Resource accessibility
- Resource Fragmentation
- Resource responsiveness
- Degree of reworking
- Target complexity
- Target memory constraints
- Target timing constraints
- Team Capability
- Development Experience
- Application experience
CONCLUSIONS

- FEW SYSTEMS COMPRISE LARGE PERCENTAGE OF COST
- MODELS RANDOMLY INACCURATE FOR DATA SET AS A WHOLE
  - DYNAMIC ENVIRONMENT
  - VARIETY OF SUPPLIERS AND FUNCTIONS
  - LACK OF EXPERIENCED SOFTWARE MODELLERS
- EXPERTISE, TOOLS, AND BASELINING CAN INCREASE MODEL ACCURACY

LESSONS LEARNED

- START DATA COLLECTION EARLY IN PROGRAM DEVELOPMENT CYCLE
- CLEARLY DEFINE EFFORT DATA - PHASES, ACTIVITIES, ORGANIZATIONS
- PROVIDE TRAINING/CONSULTATION FOR SUBJECTIVE DATA ASSESSMENT
- PROVIDE FEEDBACK OF DATA ANALYSIS TO PROGRAM ENGINEERS