A STRATEGY FOR A CREDIBLE & AUDITABLE ESTIMATION PROCESS USING ISBSG REPOSITORY

Alain Abran, Reiner Dumke, J M Desharnais, I. Ndiaye, C. Kolbe

International Workshop on Software Measurements
October 7-9, 2002, Magdeburg (Germany)
Presentation Plan

- Estimation Process
- Estimation using ISBSG
- Business Context
Estimation: IF project size is this BIG, then how much will it cost?

Software Size

Effort (staff hours)
Most common used estimation method: Expert Judgment (...or Black Magic ?)

Wishes

- Dream 1
- Dream 2
- Dream 3

Estimation - Black Box

Results - A number

? Miracle!

© 2002 Software Engineering Research Laboratory
High-level model of an estimation process

Input 1

Input 2

Input i

Estimation Model

Estimation outcomes
Presentation Plan

- Estimation Process
- Estimation using ISBSG
- Business Context
Available Repositories

- Alternatives:
  - In-house historical data sets

- Estimation tools = black boxes (in general)

- Publicly available repositories = ISBSG
  (International Software Benchmarking Standards Group)
ISBSG 1999 Data set : 749 projects

- Selection criteria:
  - with effort
  - with functional size
  - with programming languages
  - with ISBSG quality criteria

- Data set used:
  - with required info = 479 projects
  - Sub-samples by programming languages = 371
Oracle [100, 200]

\[ y = 7,7803x - 1280,7 \]

\[ R^2 = 0,3918 \]
PL1 [80, 450]  

\[ y = 8.3264x - 197.67 \]

\[ R^2 = 0.6441 \]
Powerbuilder [60, 480]

\[ y = 1.9942x + 1560.1 \]

\[ R^2 = 0.1304 \]
Telon [70, 750]

\[ y = 5.5006x + 1046.1 \]

\[ R^2 = 0.7524 \]
### Set of 12 samples - with all data points

<table>
<thead>
<tr>
<th>Programming language *</th>
<th>N</th>
<th>Functional size interval</th>
<th>N</th>
<th>Functional size interval</th>
<th>No. outliers excluded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access</td>
<td>17</td>
<td>200-1500</td>
<td>11</td>
<td>200-800</td>
<td>6</td>
</tr>
<tr>
<td>C</td>
<td>15</td>
<td>40-2500</td>
<td>9</td>
<td>200-800</td>
<td>6</td>
</tr>
<tr>
<td>C++</td>
<td>21</td>
<td>70-1500</td>
<td>12</td>
<td>70-500</td>
<td>4</td>
</tr>
<tr>
<td>Cobol</td>
<td>106</td>
<td>0-5000</td>
<td>60</td>
<td>60-400</td>
<td>14</td>
</tr>
<tr>
<td>Cobol II</td>
<td>21</td>
<td>80-2000</td>
<td>9</td>
<td>80-180</td>
<td>6</td>
</tr>
<tr>
<td>Natural</td>
<td>41</td>
<td>20-3500</td>
<td>30</td>
<td>20-620</td>
<td>2</td>
</tr>
<tr>
<td>Oracle</td>
<td>26</td>
<td>110-4300</td>
<td>19</td>
<td>100-2000</td>
<td>7</td>
</tr>
<tr>
<td>PL/1</td>
<td>29</td>
<td>80-2600</td>
<td>19</td>
<td>80-450</td>
<td>5</td>
</tr>
<tr>
<td>Powerbuilder</td>
<td>18</td>
<td>60-900</td>
<td>12</td>
<td>60-400</td>
<td>6</td>
</tr>
<tr>
<td>SQL</td>
<td>20</td>
<td>280-4400</td>
<td>11</td>
<td>280-800</td>
<td>1</td>
</tr>
<tr>
<td>Telon</td>
<td>23</td>
<td>70-1100</td>
<td>18</td>
<td>70-650</td>
<td>5</td>
</tr>
<tr>
<td>Visual Basic</td>
<td>34</td>
<td>30-2300</td>
<td>24</td>
<td>30-600</td>
<td>10</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>371</strong></td>
<td></td>
<td><strong>299</strong></td>
<td></td>
<td><strong>72</strong></td>
</tr>
</tbody>
</table>

* Programming language as recorded in ISBSG repository
<table>
<thead>
<tr>
<th>Language</th>
<th>N=377</th>
<th>Size Interval</th>
<th>Linear Regression Equation (where x = FP units)</th>
<th>R²</th>
<th>N=302</th>
<th>Size Interval</th>
<th>Linear Regression Equation (where x = FP units)</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access</td>
<td>17</td>
<td>200-1500</td>
<td>Y=-0.10x+840.8</td>
<td>0.01</td>
<td>11</td>
<td>200-800</td>
<td>Y= 0.30x + 623.5</td>
<td>0.19</td>
</tr>
<tr>
<td>C</td>
<td>15</td>
<td>40-2500</td>
<td>Y= 4.05x + 4288</td>
<td>0.19</td>
<td>9</td>
<td>200-800</td>
<td>Y= 2.34x + 2951.7</td>
<td>0.29</td>
</tr>
<tr>
<td>C++</td>
<td>21</td>
<td>70-1500</td>
<td>Y=13.43x + 1346.4</td>
<td>0.62</td>
<td>12</td>
<td>70-500</td>
<td>Y= 11.53x + 1197.1</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>750-1250</td>
<td>Y= -6.57x + 23003</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cobol</td>
<td>106</td>
<td>0-5000</td>
<td>Y=4.94x + 5269.3</td>
<td>0.36</td>
<td>60</td>
<td>60-400</td>
<td>Y= 10.83x + 299.1</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>32</td>
<td>401-3500</td>
<td>Y= 12.32x – 14.1</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cobol II</td>
<td>21</td>
<td>80-2000</td>
<td>Y= 27.80x – 3593</td>
<td>0.96</td>
<td>9</td>
<td>80-180</td>
<td>Y= 16.39x – 92.3</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td>180-500</td>
<td>Y= 26.73x – 3340.8</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Natural</td>
<td>41</td>
<td>20-3500</td>
<td>Y=10.05x – 648.9</td>
<td>0.85</td>
<td>30</td>
<td>20-620</td>
<td>Y= 6.13x + 264.9</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
<td>620-3500</td>
<td>Y= 10.53x – 1404.9</td>
<td>0.74</td>
</tr>
<tr>
<td>Oracle</td>
<td>26</td>
<td>110-4300</td>
<td>Y=6.20x + 509.7</td>
<td>0.42</td>
<td>19</td>
<td>100-2000</td>
<td>Y= 7.78x – 1280.7</td>
<td>0.39</td>
</tr>
</tbody>
</table>

© 2002 Software Engineering Research Laboratory
Table 4: Performance of ISBSG regression models on samples without outliers

<table>
<thead>
<tr>
<th>Programming languages and size range</th>
<th>RRMS(%)</th>
<th>PRED(0.25)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access [200,800]</td>
<td>15</td>
<td>91</td>
</tr>
<tr>
<td>C [200, 800]</td>
<td>50</td>
<td>22</td>
</tr>
<tr>
<td>C++ [70, 500]</td>
<td>86</td>
<td>25</td>
</tr>
<tr>
<td>C++ [750, 1250]</td>
<td>24</td>
<td>60</td>
</tr>
<tr>
<td>Cobol [60, 400]</td>
<td>42</td>
<td>43</td>
</tr>
<tr>
<td>Cobol [401, 3500]</td>
<td>51</td>
<td>35</td>
</tr>
<tr>
<td>Cobol [80, 180]</td>
<td>29</td>
<td>78</td>
</tr>
<tr>
<td>Cobol II [181, 500]</td>
<td>46</td>
<td>33</td>
</tr>
<tr>
<td>Natural [20, 620]</td>
<td>50</td>
<td>27</td>
</tr>
<tr>
<td>Natural [621, 3500]</td>
<td>35</td>
<td>33</td>
</tr>
<tr>
<td>Oracle [100, 2000]</td>
<td>120</td>
<td>21</td>
</tr>
<tr>
<td>PL1 [80, 450]</td>
<td>45</td>
<td>42</td>
</tr>
<tr>
<td>PL1 [451, 2550]</td>
<td>21</td>
<td>60</td>
</tr>
<tr>
<td>Powerbuilder [60, 480]</td>
<td>29</td>
<td>58</td>
</tr>
<tr>
<td>SQL [280, 800]</td>
<td>81</td>
<td>27</td>
</tr>
<tr>
<td>SQL [801, 4500]</td>
<td>45</td>
<td>25</td>
</tr>
<tr>
<td>Telon [70, 650]</td>
<td>22</td>
<td>56</td>
</tr>
</tbody>
</table>
Web-based COSMIC Data in next ISBSG Release
Web-based COSMIC Data (next ISBSG Release)
Regression Model - 1 Variable

![Regression Model Graph]

Cfsu vs Effort graph showing a linear trend.
Web-based COSMIC Data (next ISBSG Release)
Regression Model - 2 Variables

\[
y = 0.65 \times \text{Cfsu} + 8.9 \quad \text{R}^2 = 0.83
\]

\[
y = 0.81 \times \text{Cfsu} + 172.51 \quad \text{R}^2 = 0.57
\]
Presentation Plan

- Estimation Process
- Estimation using ISBSG
- Business Context
Key components of an estimation process

Estimation Process

Measures

Resources

Processes

Products

Audit Report

Reliability Report

Simulation Results

Productivity Simulation Model

Uncertainty Factors:
High-level Project Assumptions
  e.g.: Key Project Drivers

Risk Assessment

Confidence Report

Adjustment Process

Estimation Outcomes

© 2002 Software Engineering Research Laboratory
Estimation Process

Project Measures

- Resource
- Process
- Product

Project Context

- Project Assessment
- Productivity Database
- Estimation Database

Audit Report

Productivity Simulation Model

Simulation Results

Reliability Report

Uncertainty Factors:
- High level Project Assumptions
  - e.g.: Key Project Drivers

Risks

Adjustment Process

Software Estimation Outcomes

Confidence Report

Laboratory
Business Context

- The technical estimate based on:
  - From the estimation process with its statistical model data repository and (+ audit reports)

- The final estimate = a **business** estimate to
  - Make money
  - Minimize the risks
  - Win the business
Distinct Roles

- Technical responsibilities
  - related techniques & auditable requirements

- Business responsibilities
  - Related risks taken & documented assumptions on control variables & mitigation of risks and uncertainty

- Credibility requires both auditability and traceability
Thank you
Set C: Maintenance Requests

Effort = 2.92 FFP

Silva & Abran, 2000
Set C: Maintenance Requests

\[ y = 0.6075x + 91.289 \]

\[ R^2 = 0.1196 \]

Silva & Abran, 2000
### Set C: Maintenance Requests

<table>
<thead>
<tr>
<th>Equation</th>
<th>Model Type</th>
<th>N</th>
<th>A</th>
<th>B</th>
<th>R</th>
<th>R^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y = A + B \cdot X</td>
<td>Linear</td>
<td>19</td>
<td>91.289</td>
<td>0.607</td>
<td>0.35</td>
<td>0.12</td>
</tr>
<tr>
<td>Y = A \cdot X^B</td>
<td>Power</td>
<td>19</td>
<td>43.808</td>
<td>0.245</td>
<td>0.50</td>
<td>0.245</td>
</tr>
<tr>
<td>Y = A \cdot e^{(B \cdot X)}</td>
<td>Exponential</td>
<td>19</td>
<td>63.067</td>
<td>0.006</td>
<td>0.39</td>
<td>0.15</td>
</tr>
<tr>
<td>Y = A + B \cdot \ln(X)</td>
<td>Logarithmic</td>
<td>19</td>
<td>44.121</td>
<td>29.29</td>
<td>0.51</td>
<td>0.26</td>
</tr>
<tr>
<td>Y = A + B \cdot X</td>
<td>Hyperbolic 1</td>
<td>19</td>
<td>132.463</td>
<td>-48.330</td>
<td>0.32</td>
<td>0.10</td>
</tr>
<tr>
<td>Y = 1 / (A + B \cdot X)</td>
<td>Hyperbolic 2</td>
<td>19</td>
<td>0.0022</td>
<td>-8.8E-05</td>
<td>0.31</td>
<td>0.09</td>
</tr>
<tr>
<td>Y = X / (A + B \cdot X)</td>
<td>Hyperbolic 3</td>
<td>19</td>
<td>0.0007</td>
<td>0.016</td>
<td>0.28</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Silva & Abran, 2000
Set C: Maintenance Requests

Effort = 0.92 FFP + 126.12 Difficulty +25.9 and $R^2: 0.46$ and n = 19

If Difficulty = 0 → Effort = 0.92 FFP + 25.9
IF Difficulty = 1 → Effort = 0.92 FFP + 152.02
Set C: Maintenance Requests
Multiplicative model with size and complexity

\[ y = 0.6396x + 41.005 \]
\[ R^2 = 0.4706 \]

\[ y = 4.4917x + 82.949 \]
\[ R^2 = 0.7787 \]

Silva & Abran, 2000