Evaluation of SLIM Estimation Model Using ISBSG Repository

IWSM 2001,
Montreal, August 28-29, 2001

IPHIGÉNIE M.A. NDIAYE
A. Abran, G. Lévesque
Agenda

- Introduction
- Project’s definition
- Project’s planning
- Operation
- Interpretation
- Conclusion
Introduction

• Duration and software cost estimation => major elements in software projects planning

  ❖ Accurate estimation of both of them is really important

  ❖ => reliable estimation tool: Is SLIM reliable?
Project’s definition

• 1. Motivation
  ➢ Decision making is not an easy task
  ➢ Lot of unknowns at the beginning of the project development
  [Abran et Robillard, 1993]

• Improve the quality of decision-making of the software projects managers.
Project’s definition

2. Domain: MIS (Management information systems):
   - Software development
   - Software cost estimation models

3. Object: SLIM tool:
   - Based on Putnam’s estimation model (1978)
   - Based on the Rayleigh curve

\[
K = \left( \frac{LOC}{(C \times t^{4/3})} \right) \times 3
\]
Project’s definition

  - Projects collected from 20 countries: 35% Australia, 34.4% North America, 29.2% Europe, 0.4 South America, 1% no identified.

• Projects mostly from domain of business application: 43% IS, 33% transaction processing applications, and 5% real-time related applications
Project’s planning

• Directs criteria :
  ➢ Error analysis (Conte et al, 1986)
    ➢ Magnitude relative error
      \[ MRE = |RE| = \frac{E_{act} - E_{est}}{E_{act}} \]
    ➢ Mean magnitude of relative error \( MMRE = \frac{1}{n} \sum_{i=1}^{n} MRE_i \)
    ➢ Relative mean square error
      \[ RMS = (\overline{SE})^{1/2} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (E_{act_i} - E_{est_i})^2} \]
    ➢ Square root of mean relative error
    ➢ Prediction level
      \[ PRED(l) = \frac{k}{n} \]
  ➢ Linear regression to measure the correlation between estimated effort and real effort

NDIAYE, Abran, Lévesque  IWSM 2001
Project’s planning

• Indirect criteria : Basis criteria of ISBSG-1999 (sample) :
  ➢ No doubt about the quality of the data point :
    ❖ each project has a quality tag assigned by ISBSG, based on whether or not the data received has fully met their data collection quality requirements, that is do the consider any specific data as fully credible
  ➢ The project effort (in person-hours) is available and must be equal or greater than 400 p-h;
  ➢ The project duration (n calendar month) is available
  ➢ The programming language is available
## Project’s planning

<table>
<thead>
<tr>
<th></th>
<th>Duration (month)</th>
<th>Effort (person-hour)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of projects</td>
<td>497</td>
<td>497</td>
</tr>
<tr>
<td>Maximum</td>
<td>1</td>
<td>400</td>
</tr>
<tr>
<td>Minimum</td>
<td>84</td>
<td>138883</td>
</tr>
<tr>
<td>Average</td>
<td>10,5</td>
<td>6949</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>9,2</td>
<td>13107</td>
</tr>
<tr>
<td>Median</td>
<td>8</td>
<td>2680</td>
</tr>
</tbody>
</table>

NDIAYE, Abran, Lévesque IWSM 2001
Operation

- Natural with all 41 projects:
  - $Y = 10.05X - 648$
  - $R^2 = 0.85$
Operation

- Natural [20, 620]
- $Y = 6.13X + 264$
- $R^2 = 0.47$
Operation

- Natural [621, 3500]:
  - $Y = 10.53X - 1404$
  - $R^2 = 0.74$
Results and analysis

- Models without outliers
  - Real effort and SLIM
    - Natural [20, 620]: 60% under-estimated
    - Natural [621, 3500]: 56% over-estimated
    - MRE Natural [20, 620]: 248%
    - MRE Natural [621, 3500]: 850%

- Is SLIM a good model?
Based on the size, the duration and the language as parameters, SLIM’s estimations are very far from real effort.
### Operation

**Results and analysis**

- Real effort and SLIM:

<table>
<thead>
<tr>
<th>Types of languages</th>
<th>Number of models</th>
<th>% underestimated</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 GL</td>
<td>11</td>
<td>64</td>
</tr>
<tr>
<td>4 GL</td>
<td>17</td>
<td>88</td>
</tr>
<tr>
<td>APG</td>
<td>2</td>
<td>100</td>
</tr>
</tbody>
</table>
• SLIM’s estimations are often less than real effort

• Those Results look like Kemerer’s (1987):
  ➢ % Mean error of SLIM’s estimations = 772%
    with a minimum error of 21%

➢ SLIM has been developed with the data of project of the department of american defense and claims to be now based on +7000 projects.
Real effort and ISBSG:

- Natural [20, 620]: 70% of 30 projects are overestimated
- Natural [621, 3500]: underestimated at 56%

<table>
<thead>
<tr>
<th>Types of languages</th>
<th>Number of models</th>
<th>% underestimated</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 GL</td>
<td>11</td>
<td>55</td>
</tr>
<tr>
<td>4 GL</td>
<td>17</td>
<td>41</td>
</tr>
<tr>
<td>APG</td>
<td>2</td>
<td>50</td>
</tr>
</tbody>
</table>
Results and analysis

• SLIM et ISBSG:
  - Natural [20, 620]: ISBSG > SLIM
  - Natural [621, 3500]: ISBSG < SLIM

<table>
<thead>
<tr>
<th>Types of language</th>
<th>Number of models</th>
<th>ISBSG &gt; SLIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 GL</td>
<td>11</td>
<td>73</td>
</tr>
<tr>
<td>4 GL</td>
<td>17</td>
<td>88</td>
</tr>
<tr>
<td>APG</td>
<td>2</td>
<td>100</td>
</tr>
</tbody>
</table>

Mean relative error (MRE):
- Not many languages have a MRE of ISBSG > SLIM
- => ISBSG more reliable than SLIM
Operation

Results and analysis

• MRE:
  - 23/30 of sets of projects of the 3 types of languages have ISBSG’s estimation more accurate than SLIM’s
    - SLIM is the best only for others 4 GL [110, 950], Cobol II [181, 500], SQL
  - ISBSG is the best for others 3 GL, Access, Easytrieve, SQL Windows, APG, Telon
  - Aside those and others 4GL [110, 950], neither SLIM, nor ISBSG is acceptable for projects’ effort estimation
Operation

Results and analysis

• Square root of mean relative error (RRMS) and prediction level ($PRED(l)$):

  ➢ Models without outliers:
    ❖ RRMS: SLIM > ISBSG
    ❖ PRED (0,25): SLIM < ISBSG
    ❖ RRMS high except for « other 3 GL »
• Models without outliers
  ➢ ISBSG’s correlation coefficient always higher than SLIM’s

• Link between real effort and the one estimated by ISBSG is more accentuated than the link between the same real effort and SLIM’s estimation
Operation

Results and analysis

Models without outliers:

• $R^2 > 50\%$
  - SLIM: 25% (4GL), 20% (3GL), 0% (APG)
  - ISBSG: 50% (4GL), 56% (3GL), 50% (APG)

• $R^2 > 70\%$
  - SLIM: 6.25% (4GL), 11% (3GL), 0% (APG)
  - ISBSG: 25% (4GL), 22% (3GL), 50% (APG)
INTERPRETATION

• Interpretation context:
  ➢ Only SLIM-estimate have been used
  ➢ Use of analytic and statistical models
  ➢ Goal reached
  ➢ Field of posteriori productivity models evaluation
INTERPRETATION

- Extrapolation: Sample representativeness
  - Positive:
    - Large database (789 projects) => 457 projects
    - International projects
    - Completed projects
    - Various languages
  - Negative:
    - Not enough projects for some languages
INTERPRETATION

• Further Work:
  
  ➢ Calibrate SLIM to the development environment of a specific project before estimating.
  
  ➢ Adjust the gearing factor at the projects of the enterprise which is using SLIM, in order to adapt it at each enterprise context.
  
  ➢ Using more than one tool (model) may be a possibility of estimations improvement.
CONCLUSION

« Adding man-power to a late software projects makes it later » [Brooks, 1975]. So, we must:

- Have a good cost and duration estimation
- Have a reliable estimation model

- But SLIM isn’t eligible at this criteria of a good model in software engineering:
  
  «a productivity model is considered good, if its MRE is between ± 25% for 75% of the observations». 

NDIAYE, Abran, Lévesque  IWSM 2001
References


References


Questions and comments

?