Software Estimation: Practical Insights & Orphean Research Issues

Alain Abran
École de Technologie Supérieure, University of Québec, Montréal, Canada
alain.abran@etsmtl.ca

9th International Conference Software Engineering Advances
October 12-16, 2014 - Nice (France)
Alain Abran

20 years

- Development
- Maintenance
- Process Improvements

20 years

+ 35 PhD

ISO:
19761, 9216, 25000, 15939, 14143, 19759

© Copyrights Abran 2014
List of topics

1. Estimation: Craft or Engineering?

2. The estimation phases

3. Economic concepts for estimation models

4. Orphean research issues
List of topics

1. Estimation: Craft or Engineering?

2. The estimation phases

3. Economic concepts for estimation models

4. Orphean research issues
(Software) Estimation

Or?
Estimation expectations
Figure 1.1 Common view of an estimation process.
Figure 1.2 Some poor estimation practices observed in industry.
Figure 1.2 Some poor estimation practices observed in industry.
Figure 1.3 Some of the worst estimation practices.
Figure 1.3 Some of the worst estimation practices.
Imprecise Inputs at Feasibility Analysis – Much Greater Error Range

Margin of error = orders of magnitude!
Figure 1.2 Some poor estimation practices observed in industry.
A look at the most-known estimation approach:

The ‘COCOMO-like’ approach with its ‘cost drivers’ where:

**Effort = F(Size, +15 ‘cost drivers’)**
Figure 7.8 A step-function estimation model with irregular intervals.
Figure 7.8 A step-function estimation model with irregular intervals.
Figure 7.9 Approximation of step-function productivity models with irregular intervals.
Each COCOMO cost driver = an estimation sub-model with unknown quality & large errors.

Figure 7.9 Approximation of step-function productivity with irregular intervals.
COCOMO-like estimation models:
Effort is a function of (Size & +15 step-functions)
COCOMO-like estimation models: Effort is a function of (Size & +15 step-functions)
COCOMO-like estimation models:
Effort is a function of (Size & +15 step-functions) of unknown quality combined into a single number!

Greater Error Ranges

© Copyrights Abran 2014
COCOMO-like estimation models: Effort is a function of (Size & +15 step-functions) of unknown quality combined into a single number!

Built-in Systematic Errors & Error Propagation

© Copyrights Abran 2014
Kemerer 1987 on COCOMO81

Small scale replication study - 17 projects

<table>
<thead>
<tr>
<th></th>
<th>Basic Exponential on Size</th>
<th>Intermediate &amp; 15 cost drivers</th>
<th>Detailed &amp; 4 project phases</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>$R^2$</strong> (max=1.0)</td>
<td>0.68</td>
<td>0.60</td>
<td>0.52</td>
</tr>
<tr>
<td><strong>MMRE</strong> (mean magnitude of relative errors)</td>
<td>610%</td>
<td>583%</td>
<td>607%</td>
</tr>
</tbody>
</table>
Figure 7.2 Desired impact.
Figure 7.3 Plausible Greater Impact of Adjustments to Estimates.
Figure 7.6 Estimation - The Bundle Approach.
Figure 7.6 Estimation - The Bundle Approach.
Estimation Maths status: The search for gold!
Another Estimation Model:

- With complex mathematical formula
- Claims of being based on +4,000 projects

Still being marketed in 2014
...at a very high cost!
KEMERER 1987 on this other estimation model

Small scale replication study – 17 projects

$\text{MMRE} = 772\%$

With both large + & -
(i.e. cannot be calibrated!)
### Larger scale replication study - MMRE

<table>
<thead>
<tr>
<th>Programming language, size range [in Function Points]</th>
<th>(1) Vendor’s black-box estimation tool (%)</th>
<th>(2) White-box models built directly from the data (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access [200,800]</td>
<td>341</td>
<td>15</td>
</tr>
<tr>
<td>C [200, 800]</td>
<td>1653</td>
<td>50</td>
</tr>
<tr>
<td>C++ [70, 500]</td>
<td>97</td>
<td>86</td>
</tr>
<tr>
<td>C++ [750, 1250]</td>
<td>95</td>
<td>24</td>
</tr>
<tr>
<td>Cobol [60, 400]</td>
<td>400</td>
<td>42</td>
</tr>
<tr>
<td>Cobol [401, 3500]</td>
<td>348</td>
<td>51</td>
</tr>
<tr>
<td>Cobol II [80, 180]</td>
<td>89</td>
<td>29</td>
</tr>
<tr>
<td>Cobol II [180, 500]</td>
<td>109</td>
<td>46</td>
</tr>
<tr>
<td>Natural [20, 620]</td>
<td>243</td>
<td>50</td>
</tr>
<tr>
<td>Natural [621, 3500]</td>
<td>347</td>
<td>35</td>
</tr>
<tr>
<td>Oracle [100, 2000]</td>
<td>319</td>
<td>120</td>
</tr>
<tr>
<td>PL1 [80, 450]</td>
<td>274</td>
<td>45</td>
</tr>
<tr>
<td>PL1 [550, 2550]</td>
<td>895</td>
<td>21</td>
</tr>
<tr>
<td>Powerbuilder [60, 400]</td>
<td>95</td>
<td>29</td>
</tr>
<tr>
<td>SQL [280, 800]</td>
<td>136</td>
<td>81</td>
</tr>
<tr>
<td>SQL [801, 4500]</td>
<td>127</td>
<td>45</td>
</tr>
<tr>
<td>Telon [70, 650]</td>
<td>100</td>
<td>22</td>
</tr>
<tr>
<td>Visual Basic [30, 600]</td>
<td>122</td>
<td>54</td>
</tr>
<tr>
<td>Min</td>
<td>89</td>
<td>15</td>
</tr>
<tr>
<td>Max</td>
<td>1,653</td>
<td>120</td>
</tr>
</tbody>
</table>

© Copyrights Abran 2014
## Larger scale replication study - MMRE

<table>
<thead>
<tr>
<th>Programming language, size range [in Function Points]</th>
<th>(1) Vendor's black-box estimation tool (%)</th>
<th>(2) White-box models built directly from the data (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access [200,800]</td>
<td>341</td>
<td>15</td>
</tr>
<tr>
<td>C [200, 800]</td>
<td>1653</td>
<td>50</td>
</tr>
<tr>
<td>C++ [70, 500]</td>
<td>97</td>
<td>86</td>
</tr>
<tr>
<td>C++ [750, 1250]</td>
<td>95</td>
<td>24</td>
</tr>
<tr>
<td>Cobol [60, 400]</td>
<td>400</td>
<td>42</td>
</tr>
<tr>
<td>Cobol [401, 3500]</td>
<td>348</td>
<td>51</td>
</tr>
<tr>
<td>Cobol II [80, 180]</td>
<td>89</td>
<td>29</td>
</tr>
<tr>
<td>Cobol II [180, 500]</td>
<td>109</td>
<td>46</td>
</tr>
<tr>
<td>Natural [20, 620]</td>
<td>243</td>
<td>50</td>
</tr>
<tr>
<td>Natural [621, 3500]</td>
<td>347</td>
<td>35</td>
</tr>
<tr>
<td>Oracle [100, 2000]</td>
<td>319</td>
<td>120</td>
</tr>
<tr>
<td>PL1 [80, 450]</td>
<td>274</td>
<td>45</td>
</tr>
<tr>
<td>PL1 [550, 2550]</td>
<td>895</td>
<td>21</td>
</tr>
<tr>
<td>Powerbuilder [60, 400]</td>
<td>95</td>
<td>29</td>
</tr>
<tr>
<td>SQL [280, 800]</td>
<td>136</td>
<td>81</td>
</tr>
<tr>
<td>SQL [801, 4500]</td>
<td>127</td>
<td>45</td>
</tr>
<tr>
<td>Telon [70, 650]</td>
<td>100</td>
<td>22</td>
</tr>
<tr>
<td>Visual Basic [30, 600]</td>
<td>122</td>
<td>54</td>
</tr>
<tr>
<td>Min</td>
<td>89</td>
<td>15</td>
</tr>
<tr>
<td>Max</td>
<td>1,653</td>
<td>120</td>
</tr>
</tbody>
</table>
Another Key Issue: The quality of the inputs to the estimation models
Figure 2.5 A Normal distribution and the standards deviations.
Figure 6.2 Confidence Intervals & Sample Intervals.
Figure 5.4 Frequency distribution of the size (independent variable) in Table 5.1 with N = 212.
A New Software Metric to Complement Function Points

The Software Non-functional Assessment Process (SNAP)
Author  Very strong relationship of SNAP with Effort

\[ R^2 = 0.89 \quad (R^2 \text{ max} = 1.0) \]
Author’s assertion on Figure 4:

- $R^2 = .89$ Significance $F = 1.7 \times 10^{-23}$ Spearman = .85
- Runs = pass

&

- Spearman test for rank correlation of .85, with an associated confidence of statistical significance of greater than 99% (p-value < .0001).
Author’s assertion on Figure 4:

- $R^2 = .89$ Significance $F = 1.7 \times 10^{-23}$ Spearman = .85
- Runs = pass

&

- Spearman test for rank correlation of .85, with an associated confidence of statistical significance of greater than 99% ($p$-value < .0001).

But:

These numbers and stats are **invalid**

the necessary requirements for a regression are not met!

Presence of large outliers which distorts all stats numbers

Meaningless!!
SNAP POINTS -- FINAL BETA TEST RESULTS

- Work Effort vs. SNAP Points
- The graph shows a positive correlation between SNAP Points and Work Effort, indicated by the upward trend of the data points along the diagonal line.
What it really looked like for the range for which there is enough data points

Approximimatively:
An $R^2 = 0.3$
Not $R^2 = 0.89$
($R^2_{\text{max}} = 1.0$)

CONCLUSION: invalid approach to empirically adopt SNAP!
Hell is paved all over with good intentions!
Estimation is always urgent

Is there a ‘quick’ solution?
Estimation is always urgent

Is there a ‘quick’ solution?

The web!
Estimation is always urgent

Is ‘quick’ the right question?

The web!
Estimation is always **urgent**

Isn’t it ‘**quality**’ of the estimation model the right question?

The web!
Estimation is always urgent

Isn’t ‘**quality**’ of the estimation model the right question?

The web!
Estimation Expected Outcomes

Quick & Easy…
Estimation Outcomes!

Quick & Easy…
COCOMO-like estimation models

The ‘feel-good’ dead end!

Quick & Easy…
List of topics

1. Estimation: Craft or Engineering?

2. The estimation phases

3. Economics concepts for estimation models (fixed-variable costs, economies of scale...)

4. Orphean research issues
Figure 1.5 Uncertainty decreases over time.
Figure 1.10 Phase A:
Figure 1.10 Phase A:
Models Built with completed projects

Figure 1.8 The context of a productivity model.

© Copyrights Abran 2014
The inputs to productivity models have little uncertainty = Known Facts
Imprecise Inputs at Feasibility Analysis – Much Greater Error Range

Margin of error = orders of magnitude!
Project Scope = ?

Stakeholders initial wishes

The dreamer

Marketing

The visionnary

Accounting
Project Scope: Detailed & Approved

Stakeholders initial wishes

The dreamer

Marketing

The visionnary

Accounting

Agreed Project Scope!
Estimation Models: The Uncertainty Cone: Requirements Specs

Variability (%)

Feasibility Study  Requirements Specification  Software Development

Margin of error = orders of magnitude!

Time

© Copyrights Abran 2014
Figure 1.10 Phase B : Execution of the productivity model.
Figure 3.1 Best & Worst Case Scenarios.
Figure 3.3 Best & Worst Scenarios & Size Uncertainty.
Figure 1.11 Phase C: The adjustment process.
Figure 3.2 Most Likely Scenario & Over-optimism.
Figure 3.4 Probability Distribution of Scenarios.
Figure 7.4 Project budget = contingency = price - Optimistic scenario.
Figure 7.5 Project budget = contingency = price - Pessimistic scenario.
Figure 1.12 Phase D: Budgeting decision.
Figure 1.13 Phase E : Re-Estimation.
Figure 1.14 Phase F: Estimation Feedback Loop.
Estimator role: Provide Information
Manager role: Pick a number & Manage Risk
List of topics

1. Estimation: Craft or Engineering?

2. The estimation phases

3. Economic concepts for estimation models

4. Orphean research issues
Figure 2.16 An homogeneous size-effort dataset in software engineering.
Figure 2.17 The TELON dataset in the ISBSG 1999 Release (Abran, Ndiaye, Bourque, 2007)
Figure 2.11 Model with a fixed and variable costs.
Figure 2.15 Wedge-shaped dataset in software engineering.
Diseconomies of scale

Figure 2.19 Wedge shape with 3 data subsets with economies/diseconomies of scale.
Projects from a financial governmental organization
Projects from a financial governmental organization

Figure 12.4 The organization’s production model.
Figure 12.5  The two subsets of projects within the single dataset.
Figure 12.7  Most productive projects.
Causes: Schedule compression, users changing their minds, integrated applications…

Figure 12.6 Least productive projects.
Which estimation model to use in which contexts?

A Management Decision!

Figure 12.5  The two subsets of projects within the single dataset.
Estimation Models based on economic concepts

A large scale success story:

- Embedded software domain

Requirements & Specifications:
  - In-house
  - Model driven using Matlab-Simulink

Software development:
  - Outsourced across the world
  - with qualified suppliers
Estimation Models based on economic concepts

Initial productivity models developed with 20 to 30 projects for each software supplier:

- Based on 2nd generation COSMIC size method
- $R^2$ within the 0.8 to 0.9+ range
- MMRE varies for each supplier
- Info on both fixed & variable costs used to compare suppliers:
  - Simple models that ‘talk’ to managers based on international standards – No ‘black boxes’ & game playing with numbers!
- Info on variance to negotiate next projects
Automated COSMIC measurement

- + 300 projects to size and estimate each at rush time every year.
- Investment in automation of functional size measurement (with a PhD student):
  - Automation results verified with duplicate measurements over +70 projects (manual & automated).
  - Accuracy of size automation:
    - Prototype: 96%
    - Final automation tool: 99+%
Other usages of functional size measurement

- Prediction model of memory size based on the size of the functional specifications

- Balancing the workload within the team of 100 engineers preparing the detailed software specifications for outsourcing

- Setting annual productivity increases to their network of software suppliers
  - as mandated to their hardware suppliers
Lessons learned

This organization:

- did not look for miracles (quick & at no cost)!

- They invested time & monies to build a competitive advantage by:
  - Collecting historical data
  - Using standards for measurement
  - Developing minimum statistical skills
  - Being transparent with software suppliers
List of topics

1. Estimation: Craft or Engineering?

2. The estimation phases

3. Economic concepts for estimation models

4. Orphean research issues
Most research focus here

- Product
- Process
- Constraints

Repository of productivity data
Repository of estimates

Confidence in the inputs
Confidence in the estimation model
Initial range of estimates
Assumptions, risks & other cost drivers

Adjustment process

INFORMATION
Expected ranges of estimates values

A unique project budget
Contingency funds

Business risks

© Copyrights Abran 2014
Most papers focus here

…and search for ‘accurate’ estimate
...and many bundle all factors into ‘Black boxes’

...while searching for ‘accurate’ estimate
Simple models easy to understand by management
Issues on Software Measurement & Estimation Inputs

- Product
- Process
- Constraints

Confidence in the inputs
Confidence in the estimation model
Initial range of estimates
Assumptions, risks & other cost drivers
Adjustment process

Estimation process improvements
Re-estimation
Project monitoring & control

INFORMATION Expected ranges of estimates values

Business risks

A unique project budget
Contingency funds

© Copyrights Abran 2014
Software Size?

- Lines of code

Or

- Function Points:
  - +30 variations
  - & 5 International Standards!
Figure 2.7 The kurtosis in a normal distribution.
Turning dust into gold…
FP to LOC conversion ratios in Estimation Models

What happened to Ariane 5 spacecraft ... and why?
Issues on Estimation Inputs

Which method for software functional size:

- **1st generation: IFPUG Function Points - 1979**
  - Innovator (in 1979, but not in 2004!)
  - Systematic errors! (step function with min & max)
  - Invalid maths!
  - No measurement unit!
  - Still cannot be automated & be compliant after 35 years!

- **2nd generation: COSMIC Function Points - 2003**
  - Strengths based on metrology principles
  - Can be automated & compliant to ISO
  - Applicable across domains
  - Free & + 15 translations
Other Issues on Estimation Inputs

- Unsound sizing methods compounding mistakes:
  - Usecase Points
  - Story Points

- For incomplete software requirements documents, lack of independently verified approximate sizing method
Estimation Models: The Uncertainty Cone: Requirements Specs
Scales in Plans - Architects & Engineers
Scales in Software Documents?

Requirements Engineer

Collect & document software requirements

SRS Document (with unknown quality issues)

Software stakeholders

© Almakadmeh 2013

© Copyrights Abran 2014
Scales in Software Requirements Texts?

1. Requirements Engineer
2. Collect & document software requirements
3. SRS Document (with unknown quality issues)
4. Approximation techniques
5. Functional size approximation (unknown quality of approximation)

© Almakadmeh 2013

© Copyrights Abran 2014
A functional size approximation technique: Unknown Performance..!

Early & Quick COSMIC technique

1. Decomposition of functional components in hierarchy
   + definition of functional components in the hierarchy
2. Table of approximate size intervals (Black-box)

Approximator

Approximation of functional size

Functional size approximation (with unknown evaluation of reproducibility & accuracy)

1. Inexistence of procedural guidelines for decomposition & identification of granularity levels
2. No documentation about approximate size intervals

© Almakadmeh 2013

© Copyrights Abran 2014
An investigation of an existing functional size approximation technique: reproducibility

### Difference of functional size approximation

<table>
<thead>
<tr>
<th>Participant code</th>
<th>Approximate functional size using the E&amp;Q COSMIC technique (Min, Most-likely, Max) (in CFP)</th>
<th>Percentage difference in functional size approximation (w.r.t. Most-likely value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A6</td>
<td>(45, 74, 93)</td>
<td>− 90%</td>
</tr>
<tr>
<td>A12</td>
<td>(57, 114, 179)</td>
<td>− 84%</td>
</tr>
<tr>
<td>A3</td>
<td>(238, 543, 910)</td>
<td>− 23%</td>
</tr>
<tr>
<td>A9</td>
<td>(250, 545, 909)</td>
<td>− 23%</td>
</tr>
<tr>
<td>A5</td>
<td>(299, 592, 962)</td>
<td>− 16%</td>
</tr>
<tr>
<td>A2</td>
<td>(250, 705, 1250)</td>
<td>0%</td>
</tr>
<tr>
<td>A1</td>
<td>(521, 1071, 1616)</td>
<td>+ 52%</td>
</tr>
<tr>
<td>A11</td>
<td>(581, 1185, 1972)</td>
<td>+ 68%</td>
</tr>
<tr>
<td>A8</td>
<td>(697, 1454, 2472)</td>
<td>+ 106%</td>
</tr>
<tr>
<td>A7</td>
<td>(964, 2077, 3450)</td>
<td>+ 195%</td>
</tr>
<tr>
<td>A4</td>
<td>(1181, 2369, 3957)</td>
<td>+ 236%</td>
</tr>
<tr>
<td>A10</td>
<td>(2265, 4510, 7408)</td>
<td>+ 540%</td>
</tr>
<tr>
<td>Minimum</td>
<td></td>
<td>− 90%</td>
</tr>
<tr>
<td>Maximum</td>
<td></td>
<td>+ 540%</td>
</tr>
</tbody>
</table>

12 Participants

© Almakadmeh 2013
An investigation of an existing functional size approximation technique: accuracy

Accuracy of the functional size approximation

<table>
<thead>
<tr>
<th>Participant code</th>
<th>Approximated functional size using the E&amp;Q COSMIC technique in CFP (min, most-likely, max) (1)</th>
<th>Reference functional size for accuracy criteria (2)</th>
<th>MRE calculated using values in (1) and (2) (min, most-likely, max)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A6</td>
<td>(45, 74, 93)</td>
<td></td>
<td>(43%, 7%, 17%)</td>
</tr>
<tr>
<td>A12</td>
<td>(57, 114, 179)</td>
<td></td>
<td>(28%, 44%, 126%)</td>
</tr>
<tr>
<td>A3</td>
<td>(238, 543, 910)</td>
<td></td>
<td>(200%, 585%, 1047%)</td>
</tr>
<tr>
<td>A9</td>
<td>(250, 545, 909)</td>
<td></td>
<td>(215%, 587%, 1046%)</td>
</tr>
<tr>
<td>A5</td>
<td>(299, 592, 962)</td>
<td></td>
<td>(277%, 646%, 1113%)</td>
</tr>
<tr>
<td>A2</td>
<td>(250, 705, 1250)</td>
<td></td>
<td>(215%, 789%, 1476%)</td>
</tr>
<tr>
<td>A1</td>
<td>(521, 1071, 1616)</td>
<td></td>
<td>(557%, 1251%, 1938%)</td>
</tr>
<tr>
<td>A11</td>
<td>(581, 1185, 1972)</td>
<td></td>
<td>(633%, 1394%, 2387%)</td>
</tr>
<tr>
<td>A8</td>
<td>(697, 1454, 2472)</td>
<td></td>
<td>(779%, 1733%, 3017%)</td>
</tr>
<tr>
<td>A7</td>
<td>(964, 2077, 3450)</td>
<td></td>
<td>(1115%, 2519%, 4250%)</td>
</tr>
<tr>
<td>A4</td>
<td>(1181, 2369, 3957)</td>
<td></td>
<td>(1389%, 2887%, 4890%)</td>
</tr>
<tr>
<td>A10</td>
<td>(2265, 4510, 7408)</td>
<td></td>
<td>(2756%, 5587%, 9241%)</td>
</tr>
<tr>
<td><strong>Average MRE on functional size approximations (all 12 participants)</strong></td>
<td></td>
<td><strong>(684%, 1502%, 2546%)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Average MRE on functional size approximations (except participants A6 &amp; A12)</strong></td>
<td></td>
<td><strong>(814%, 1798%, 3041%)</strong></td>
<td></td>
</tr>
</tbody>
</table>
Estimation Approaches

The ‘*feel-good*’ dead end!

Quick & Easy…
Building ‘good’ estimation process & good estimation models

It requires:

- Recognition of uncertainties: how to recognize this and how to deal with it
- The estimator has to provide information, not a single estimate
- The manager has to select a single budget number, and manage risks through contingency planning.
- Discipline, rigor, commitments and $$$
Figure 1.10 Phase B: Execution of the productivity model.
Orphean research issues:
Research on software estimation dates back to the early 70’s, but much still remain……

Figure 1.13 Phase E : Re-Estimation.

© Copyrights Abran 2014
Orphean research issues:
Research on software estimation dates back to the early 70’s, but much still remain......

Figure 1.13 Phase E : Re-Estimation.

© Copyrights Abran 2014
Orphean research issues:
Research on software estimation dates back to the early 70’s, but much still remain……

Figure 1.13 Phase E : Re-Estimation.
You want to know more?