Data Driven Cost Estimating

And the Role of Industry and Private Data

Karen McRitchie
Galorath Incorporated

https://itconfidence2016.wordpress.com
In the Beginning

Got Data?
Not yet.
Now

Data Driven Cost Estimating
And the Role of Industry and Private Data

![Development Effort Hours vs Effective Functions](https://itconfidence2016.wordpress.com)
Sources of Data

Public: Commercial databases, academic studies

Semipublic: Government databases

Private
### Public versus Private – Comparisons Between...

<table>
<thead>
<tr>
<th>Public Data</th>
<th>Private Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can be <strong>shared</strong></td>
<td>Must be <strong>sanitized</strong>, if shareable at all</td>
</tr>
<tr>
<td>Supposedly <strong>diverse</strong></td>
<td>Decidedly <strong>not diverse</strong></td>
</tr>
<tr>
<td>Typically “<strong>narrow</strong>” data coverage</td>
<td>Typically “<strong>wider</strong>” data coverage</td>
</tr>
<tr>
<td>Project knowledge usually general</td>
<td>Project knowledge can be <strong>intimate</strong></td>
</tr>
<tr>
<td>Several <strong>Categorical</strong> labels may fit</td>
<td><strong>Categorical</strong> labels are more specific</td>
</tr>
</tbody>
</table>
Data Gathering – Beggars Can’t Be Choosy

A List of “Nice to Haves”
Scope generally is considered most important; all else is negotiable. Though even scope can be traded away…

- Platform
- Application
- Scope
- Origin
- Effort
- Duration
- Cost
- Included labor
- Included activities
- Defects
- Staffing
- Complexity
- Extraordinary circumstances
Private Data – Bonus
Data Collection Over Time

1. **Up front:**
To gauge scope creep from start to finish. Very important for calibration, since project estimates also are made at the beginning.

2. **In process:**
During development for management, for internal use in identifying issues and gauging progress

3. **Post mortem:**
Upon completing development, this becomes the bedrock for a project repository

4. **In service:**
During maintenance to gauge life-cycle costs
Private – The “Data Under Duress” Scenario

“The Data…”
  • “...may air our dirty laundry.”
  • “...reveals our labor rates.”
  • “... is not relevant or is different from other projects”
  • “...doesn’t exist.”
  • “(...doesn’t exist, although we can’t tell you that.).”
  • “...is not clean.”
  • “...may threaten our bid strategy.”
Data Driven Cost Estimating

And the Role of Industry and Private Data

Data Sharing – Sunlight Is the Best Policy

• Your laundry isn’t so dirty. Everyone knows developing software is hard and takes more time than planned.

• You don’t need to reveal labor rates.

• Your competitors won’t ever see your data.

• You have more data than you know, even if you don’t record effort and scope.

• Let us save you time and try and clean it for you.

• Resultant data products help you bid more successfully.
Data Aggregating – A Game of Least Common Denominator

Combining data, you rely on whatever all the data sets share.

- Platform
- Scope
- Origin
- Effort
- Duration
- Included labor
- Defects
- Complexity
- Circumstances
- Platform
- Application
- Scope
- Effort
- Included labor
- Included activities

Although with opportunity for side studies.
**Task/Artifact Based Sizing - Creating Custom Metrics for Estimation**

Create custom metrics by ‘back firing’ from accounting data. We support this offline and now directly in SEER-SEM.

### Summary Metrics

<table>
<thead>
<tr>
<th>Task</th>
<th>Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>PX1-1758</td>
<td>133</td>
</tr>
<tr>
<td>PX1-3658</td>
<td>28</td>
</tr>
<tr>
<td>PX0-7883</td>
<td>57</td>
</tr>
<tr>
<td>PX1-1913</td>
<td>115</td>
</tr>
<tr>
<td>PX1-6462</td>
<td>112</td>
</tr>
<tr>
<td>PX0-4531</td>
<td>53</td>
</tr>
<tr>
<td>PX0-3619</td>
<td>210</td>
</tr>
<tr>
<td>PX1-7922</td>
<td>16</td>
</tr>
<tr>
<td>PX0-3973</td>
<td>126</td>
</tr>
<tr>
<td>PX0-4620</td>
<td>8</td>
</tr>
<tr>
<td>PX0-3311</td>
<td>192</td>
</tr>
<tr>
<td>PX0-2488</td>
<td>117</td>
</tr>
<tr>
<td>PX1-4917</td>
<td>130</td>
</tr>
<tr>
<td>PX1-3708</td>
<td>140</td>
</tr>
<tr>
<td>PX0-4318</td>
<td>159</td>
</tr>
<tr>
<td>PX0-5449</td>
<td>204</td>
</tr>
</tbody>
</table>

### Size Metric Definition

<table>
<thead>
<tr>
<th>Entry Name</th>
<th>Value</th>
<th>Units</th>
<th>Metric</th>
<th>Assumptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple Batch</td>
<td>54</td>
<td>Hours</td>
<td>3.45 FP</td>
<td></td>
</tr>
<tr>
<td>Nominal Batch</td>
<td>121.5</td>
<td>Hours</td>
<td>7.76 FP</td>
<td></td>
</tr>
<tr>
<td>Complex Batch</td>
<td>154.25</td>
<td>Hours</td>
<td>9.85 FP</td>
<td></td>
</tr>
</tbody>
</table>

### Program: Blast Processing

#### Programs

- Simple Batch
- Nominal Batch
- Complex Batch

#### Mainframe Transaction Tasks

<table>
<thead>
<tr>
<th>Entry Name</th>
<th>Value</th>
<th>Units</th>
<th>Metric</th>
<th>Assumptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple Batch</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nominal Batch</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complex Batch</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Software phase at estimate

<table>
<thead>
<tr>
<th>Software phase</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Code</td>
<td>3</td>
</tr>
</tbody>
</table>
Meta-Data

- Capture Project Patterns for estimation
  - Consistent structures make comparing projects easier
  - Avoids errors of omission in estimation
  - Facilitates data collection by identifying common categories
Dynamic Calibration – making adjustments on the fly

Project Calibration Data

Current Database Demographics

<table>
<thead>
<tr>
<th>Platform</th>
<th>Total Points In Selected Database</th>
<th>Points Matching Platform and/or Application</th>
<th>Points Matching Platform Only</th>
<th>Points Matching Application Only</th>
<th>Points Matching Platform and Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application</td>
<td>7,849</td>
<td>3,425</td>
<td>3,421</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

Calibration Selections

- Calibration Points From Class: Klose
  - 4 orig 6 removed
  - Points Manually Selected: 10
  - Calibration Points In Current Database: 14

Calibrations

- Effort: 0.99
- Schedule: 0.79

Points Used for Calibration and Available For Review:

<table>
<thead>
<tr>
<th>Name</th>
<th>Effort Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISBSG D&amp;E 10276</td>
<td>3,876</td>
</tr>
<tr>
<td>ISBSG D&amp;E 10537</td>
<td>1,197</td>
</tr>
<tr>
<td>ISBSG D&amp;E 10559</td>
<td>1,414</td>
</tr>
<tr>
<td>ISBSG D&amp;E 10570</td>
<td>3,034</td>
</tr>
<tr>
<td>ISBSG D&amp;E 10706</td>
<td>1,667</td>
</tr>
<tr>
<td>ISBSG D&amp;E 10738</td>
<td>15,840</td>
</tr>
</tbody>
</table>
Static Calibration/CER Generation

- Requires data prep and data processing
- Running the regressions is the easy part
- Be prepared to rinse and repeat
Data Driven Cost Estimating
And the Role of Industry and Private Data

Process Shared Data For Analysis

1. Data Corrections
2. Specify Records’ ‘Productivity Type’ When Missing
3. Add Back Missing Phases
4. Convert Lines To Logical Using Recommended Proportions
5. Remove Records With Identified Data Issues

6. Evaluate Records For Unidentified Data Issues
7. Remove Records Lacking Size
8. Remove Records Lacking Effort
9. Remove Records That Are Not Final Submissions
10. Remove Records That Are Not Full Development

11. Select Records Satisfying All Criteria
12. Check Outliers To Identify Suspect Records
13. Specify SEER Knowledge Bases, Notes and Descriptions
14. Determine Staffing Constraints For Modeling
15. Output to SEER Data Products

https://itconfidence2016.wordpress.com
Forensics

Schedule. Was it “stop and start”? Were there schedule constraints?

Resources. Were there hard-hitting resource constraints?

Volatility. Did requirements undergo extraordinary evolution?

Manager’s Objectives. Was it to complete the project in minimum time or at least cost?

Effort. Are effort figures actually derived from cost figures?

When creating the estimate --- adjustments for extraordinary conditions may be possible within the software estimating model.
Exploratory Data Analysis (ANOVA or regressions)

Helps to determine highest explanatory variables and functional forms.

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>ESTIMATED NAME</th>
<th>COEFFICIENT</th>
<th>T-RATIO</th>
<th>P-VALUE</th>
<th>DEFINITION OF VARIABLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR2</td>
<td>1 when resource level is 2 or above * log UFPs</td>
<td>0.2842</td>
<td>5.52</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>R2_DUM</td>
<td>1 when resource level is 2 or above</td>
<td>-1.7423</td>
<td>-6.25</td>
<td>0.032</td>
<td></td>
</tr>
<tr>
<td>LR4</td>
<td>1 when resource level is 4 * log UFPs</td>
<td>0.0436</td>
<td>2.15</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>FOURGL</td>
<td>1 when 4GL language is used</td>
<td>-0.4255</td>
<td>-8.72</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>LNUFP</td>
<td>log of UFPs</td>
<td>0.6960</td>
<td>29.36</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>PLAN</td>
<td>1 when the Plan phase is included, according to revised list of phases</td>
<td>0.6842</td>
<td>2.38</td>
<td>0.017</td>
<td></td>
</tr>
<tr>
<td>BUILD</td>
<td>1 when the Build phase is included, according to revised list of phases</td>
<td>-1.3634</td>
<td>-3.31</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>TST</td>
<td>1 when the Test phase is included, according to revised list of phases</td>
<td>1.1291</td>
<td>3.26</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>IMPL2</td>
<td>1 when the Implementation phase is included, according to ORIGINAL list</td>
<td>0.0907</td>
<td>1.68</td>
<td>0.094</td>
<td></td>
</tr>
<tr>
<td>LPLAN</td>
<td>1 for non-critical Business projects</td>
<td>-0.1511</td>
<td>-2.79</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>LSPEC</td>
<td>1 for non-critical Client-Server projects</td>
<td>0.1249</td>
<td>7.33</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>LBUILD</td>
<td>0.036 BUILD * log of UFPs</td>
<td>0.1721</td>
<td>2.10</td>
<td>0.036</td>
<td></td>
</tr>
<tr>
<td>LTEST</td>
<td>0.001 LTEST * log of UFPs</td>
<td>-0.2344</td>
<td>-3.29</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>BUSINESS</td>
<td>0 for non-critical Business projects</td>
<td>-0.2778</td>
<td>-5.87</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>CLISERV</td>
<td>0.078 1 for non-critical Client-Server projects</td>
<td>-0.1593</td>
<td>-1.76</td>
<td>0.078</td>
<td></td>
</tr>
<tr>
<td>CONSTANT</td>
<td>4.0837</td>
<td>33.85</td>
<td></td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

**Principal Components Analysis** also is useful for reducing the number of variables.

Typical forms:

\[
Effort = a \times size^{entropy}
\]

\[
Duration = b \times size^{duration\ entropy}
\]
Data Driven Cost Estimating

Classification Automation

Automated classification:
• Enables consistency
• Reduces mistakes
Typical Normalizations

• Adjustment to Line of Code or Function Point measures
  • Methods vary
  • For private data, frequently no adjustment is made – if customer prefers a given metric “flavor”, we use it

• Adjustments to re-include missing activities or labor
  • Typically using simple proportions derived from other projects in the sample

• Inferring peak staffing
  • Informed by average staffing, when known or as calculated (example on next page)
Peak Staff - Significant, Often Missing

**Staffing constraints have a significant effect on estimating**; when not specified, they should be inferred when possible:

- Use reported peak staff when available
- When not reported, based on analysis, specify peak staff at $X \times$ average staff with no constraints below $Y$ FTEs
Oversight

Data processing is like coding:
Peer reviews & crosschecks
Work needs to be tested and retested
Sanity Checks

**SEER Metrics** is used to plot data within probability bands. Points outside 2 sigma are examined.
More vs. Less Data – Double-Edged Sword

**Less Data**
- **Harder** to spot outliers
- **Easier** to explain them
- **Harder** to systemically correct

**More Data**
- **Easier** to spot outliers
- **Harder** to explain them
- **Easier** to systemically correct
Calibration Automation

Automation for solving form \( y = a \cdot s^b \)

OLS using SEER Metrics

LAE using Excel
Measuring Calibration Efficacy

Calibration efficacy can be:

- **Specific** – “Does this stratification produce enough predictions within X percent?”
- **Relative** – “Does another stratification work better while capturing the requisite estimating scenarios?”

<table>
<thead>
<tr>
<th>Stratification A</th>
<th>Effort</th>
<th>Duration</th>
<th>Stratification B</th>
<th>Effort</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median absolute deviation</td>
<td>min</td>
<td>51%</td>
<td>78%</td>
<td>Median absolute deviation</td>
<td>min</td>
</tr>
<tr>
<td>Mean absolute deviation</td>
<td>min</td>
<td>80%</td>
<td>163%</td>
<td>Mean absolute deviation</td>
<td>min</td>
</tr>
<tr>
<td>Standard error (normalized)</td>
<td>min</td>
<td>88%</td>
<td>75%</td>
<td>Standard error (normalized)</td>
<td>min</td>
</tr>
<tr>
<td>Prediction within 10%</td>
<td>max</td>
<td>6%</td>
<td>13%</td>
<td>Prediction within 10%</td>
<td>max</td>
</tr>
<tr>
<td>Prediction within 25%</td>
<td>max</td>
<td>18%</td>
<td>25%</td>
<td>Prediction within 25%</td>
<td>max</td>
</tr>
<tr>
<td>R-squared</td>
<td>max</td>
<td>66%</td>
<td>14%</td>
<td>R-squared</td>
<td>max</td>
</tr>
</tbody>
</table>

Number of records: 109

<table>
<thead>
<tr>
<th>Effort</th>
<th>Duration</th>
<th>Effort</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiplier</td>
<td>0.95</td>
<td>2.04</td>
<td>Multiplier</td>
</tr>
<tr>
<td>Exponent</td>
<td>0.84</td>
<td>0.23</td>
<td>Exponent</td>
</tr>
<tr>
<td>Original multiplier</td>
<td>4.65</td>
<td>2.19</td>
<td>Original multiplier</td>
</tr>
<tr>
<td>OLS Multiplier</td>
<td>4.66</td>
<td>1.51</td>
<td>OLS Multiplier</td>
</tr>
<tr>
<td>OLS Exponent</td>
<td>0.83</td>
<td>0.26</td>
<td>OLS Exponent</td>
</tr>
</tbody>
</table>
## Typical Statistics

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Magnitude of Relative Error (MMRE)</td>
<td>Percentage variation between actual and estimate with actual as divisor. <em>Closer to zero the better.</em></td>
</tr>
<tr>
<td>Average Estimate Ratio</td>
<td>The average of (actual / estimate) ratios. <em>Closer to one the better.</em></td>
</tr>
<tr>
<td>Median Estimate Ratio</td>
<td>The median of (actual / estimate) ratios. <em>Closer to one the better.</em></td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>The standard deviation of estimate vs. actual variances. <em>Closer to zero the better.</em></td>
</tr>
<tr>
<td>Prediction (x)</td>
<td>The percent of estimates falling within x% of actual outcomes. <em>Closer to 100% the better.</em></td>
</tr>
<tr>
<td>R-squared</td>
<td>Goodness of fit. (The goal is to purely “fit the data” and so this is a good measure.) <em>Closer to one the better.</em></td>
</tr>
</tbody>
</table>

Other statistics

Chow for structural breaks (such as may occur as scope grows), F test to detect difference between two samples (such as estimate versus actual),
Side Studies

An example of the studies possible, in addition to classic size-effort and duration-effort relationships:

1. Size growth studies, to understand estimated versus actual system size
2. The impact of staffing on project productivity
3. Determinants of project duration
4. Ability of project requirements to determine effort and duration
5. Variations in productivity by program, division/contractor, maturity rating, programming language, toolset, etc.
6. Determinants of project phases’ relative durations
7. Standardized, multiple tags for projects, yielding more descriptive information and more data for specific queries
8. Ability to predict maintenance and sustainment
9. What taxonomy has the best explanatory versus descriptive power?
Private/Local Data

- Use to shape the estimation process
  - Sizing
  - Project Patterns
  - Productivity Tuning
  - Economic Factors
  - Estimate Catalogs
- Focus is consistency, repeatability and accuracy

Public/Industry Data

- Use to target major trends
  - Calibration/CER Generation
  - Productivity Benchmarking
- Sanity Check Your Estimate
- Vendor/Competitor Evaluation
- Evaluate Industry Demographics
Applying Data To Estimation

- It’s not a lights out process
  - Requires analysts, SMEs, humans to process evaluate and identify where it can add value
- Getting started can be slow and bumpy
  - As the process is repeated, it gets easier
  - Automate as much as possible
- Revisit processes
- Measure the benefits
  - Accuracy improvements, estimate turnaround, confidence in estimates