
Carmine Gravino, Sergio Di Martino, Filomena Ferrucci, Luigi Buglione
**Goals:**

- **G1.** To receive suggestions about a statistical usage of public repository data from software projects in order to improve the selection of the proper dataset for building the estimation model.

- **G2.** To discuss the way web projects fit with different FSM methods for estimation purposes.

- **G3.** To discuss the profitable usage of cross-company datasets by a software organization.
Introduction

- Early Effort Estimation: motivations

Sizing and Estimating Web Applications

- Using COSMIC
- Using other Size Measures

Data Sets Descriptions & Research Method

- Single-company dataset
- Cross-company dataset
- Descriptive Statistics
- Research Method: Techniques & Evaluation Criteria

Obtained Results & Data Analysis

- Using OLSR
- Using CBR
- Comparing the accuracy of obtained estimates

Conclusions & Prospects
• The demand for web applications is increased
  ✓ Web apps are an essential support for the activities of business organizations

• The complexity and the size of such applications have also dramatically augmented

• Estimating the development effort/cost for a web application is crucial for the competitiveness of a company
  ✓ Good effort estimation techniques can assist project managers in planning the development and in allocating resources adequately
  ✓ ... but estimating should be obtained in a controlled, not subjective way
Early Effort Estimation: motivations

- the 5W’s+H rule
  ✓ From journalism, a common-sense accepted rule is to put in evidence for a phenomenon of interest the
    ❖ Who, What, Why, When, Where and How
  ✓ Dealing with Estimation, ‘When’ becomes a very relevant information, impacting on a series of assumptions the project will take
    ❖ Goal: balancing earlier estimates as much as possible project efforts & costs within acceptable error margins due to uncertainty
    ❖ The ‘cone of uncertainty’ is a very expressive and immediate visualization for such concepts
Introduction
- Early Effort Estimation: motivations

Sizing and Estimating Web Applications
- Using COSMIC
- Using other Size Measures

Data Sets Descriptions & Research Method
- Single-company dataset
- Cross-company dataset
- Descriptive Statistics
- Research Method: Techniques & Evaluation Criteria

Obtained Results & Data Analysis
- Using OLSR
- Using CBR
- Comparing the accuracy of obtained estimates

Conclusions & Prospects
**Sizing & Estimating Web Apps**

**Using COSMIC**

- **COSMIC** ([www.cosmicon.com](http://www.cosmicon.com))
  - 2nd generation FSM method (FSMM)
    - multi-layer, multi-perspective, no weights per BFC
  - easier to be count (data movements: E, X, R, W)
  - better fit for counting web environments than other FSMM

- **Previous contributions**
  - Rollo’s proposal (2000)
    - described the difficulties of FPA when applied to web applications and than showed that COSMIC resulted to be the most flexible approach for counting the functional size of web applications (no empirical study)
  - Mendes et al.‘s study (2002)
    - 37 hypermedia systems developed by students; OLSR; no positive results
  - Ferrucci et al.‘s study (2006)
    - 44 web applications developed by students; OLSR; encouraging results
  - Ferrucci et al.‘s study (2008)
    - 15 web applications from industry; OLSR; good estimations
**Sizing & Estimating Web Apps**

**Using other size measures**

- **Web Objects**
  - ✓ technique FP-based created by Don Reifer
  - ✓ added 4 specific BFC for web environments
    - Multimedia files, Web Building Blocks, Scripts, Links

- **Case studies**
  - ✓ Ruhe et al.’s study (2003)
    - 12 web applications from industry; OLSR; good estimations
  - ✓ Ruhe et al.’s study (2003)
    - 12 web applications from industry; Web-COBRA; good estimations
  - ✓ Ferrucci et al.’s study (2008)
    - 15 web applications from industry; OLSR; good estimations
• **OomFPWeb (Abrahão et al.)**
  - the method maps the Function Points concepts into the primitives used in the conceptual modeling phase of OOWS
  - OOWS is a method for producing software for the Web
  - A preliminary evaluation of the OOmFPWeb provided interesting results

• **Measures based on W2000 design artifacts**
  - Baresi and Morasca’s studies (2003, 2007) for identifying the attributes related to the effort for designing Web applications
    - they did not focus on the total development (effort)

• **Measures specific for Web applications such as number of Web pages, media elements, internal links, etc...**
  - Mendes et al.’s studies (2001-2008); Ferrucci et al.’s studies (2006, 2007); using Web applications developed by academic students and by companies employing several techniques, such as CBR, OLSR, and CART
To estimate:
- data from past projects, in terms of size measures that are related to effort
- a presumed size of the new project, in terms of the same measures to produce an estimation of its effort/cost development

The set of past data play a fundamental role! ... but:
- It is time-consuming to create these dataset
- Data have to be collected in a controlled and consistent way
- New companies do not have developed past projects

May a web company adopt a dataset coming from other software houses?

Mendes et al.’ studies (2004) and a replication (2008)
- TukuTuku dataset
- OLSR and the CBR as estimation techniques
- (as expected) the models obtained with the single-company dataset provided much better estimation than those obtained with the cross-company dataset
- Replication confirmed the results of the first two studies
Main problems can occur when relying on single-company data sets (from Briand et al. 2000, Mendes et al., 2007):

- The time required to accumulate enough data on past projects from a single company may be prohibitive.
- By the time the data set is large to be of use, technologies used by the company may have changed, and older projects may no longer be representative of current practices.

The use of cross-company data sets has problems of its own (from Mendes et al., 2007):

- Differences in processes and practices may result in trends that may differ significantly across companies.
- Projects should be partitioned (e.g. according to their completion dates) to identify those that used current development practices from those that did not.
Agenda

- **Introduction**
  - Early Effort Estimation: motivations

- **Sizing and Estimating Web Applications**
  - Using COSMIC
  - Using other Size Measures

- **Data Sets Descriptions & Research Method**
  - Single-company dataset
  - Cross-company dataset
  - Descriptive Statistics
  - Research Method: Techniques & Evaluation Criteria

- **Obtained Results & Data Analysis**
  - Using OLSR
  - Using CBR
  - Comparing the accuracy of obtained estimates

- **Conclusions & Prospects**
• Data for our study provided by an Italian software company
  ✓ specialized in the design, development, and management of solutions for Web portals, Geographical Information Systems, etc.
  ✓ certified ISO 9001:2000, and also a certified partner of Microsoft, Oracle, and ESRI

• The company provided us a set of 15 Web applications
  ✓ e-government, e-banking, Web portals, and Intranet applications
  ✓ developed by exploiting a wide range of Web-oriented technologies, such as J2EE, ASP.NET, etc.
• ISBSG r10 (2007)
  ✓ The most used repository with FSMM data, managed by ISBSG
  ✓ r10 contains 4106 worldwide project data
  ✓ 100+ project attributes contained
  ✓ 117 web development projects selected as a starting point

  ✓ Several steps were run for filtering the project attributes of interest

<table>
<thead>
<tr>
<th>Step</th>
<th>Attribute</th>
<th>Filter</th>
<th>Project Excluded</th>
<th>Remaining Projects</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Count Approach</td>
<td>= COSMIC-FFP</td>
<td>3989</td>
<td>117</td>
</tr>
<tr>
<td>2</td>
<td>Data Quality Rating (DQR)</td>
<td>= {A</td>
<td>B}</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>Web Development</td>
<td>= {non-blanks}</td>
<td>84</td>
<td>26</td>
</tr>
<tr>
<td>4</td>
<td>Application Type</td>
<td>= {New Development}</td>
<td>10</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>= {Enhancement}</td>
<td>20</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>= {Re-Development}</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
There are of course many differences between the two datasets

- The product functional size of the applications is quite different, with the average cross-company sizes in a 7:1 ratio against those from the single-company dataset, in terms of person/hours
- Productivity
  - [0.1 – 2.64] for cross-company dataset (excluding the outlier with 40.9)
  - [1.09 – 2.36] for single-company dataset
• **OLSR (Ordinary Least Square Regression)**
  ✓ it explores the relationship between a dependent variable and one or more independent variables, providing a prediction model

  ✓ typical equation: $y = b_1x_1 + b_2x_2 + ... + b_nx_n + c$

  ✓ typical indicators used for evaluating the quality of resulting models
    ❖ $R^2$, F value, p-value, t-value

  ✓ in this study OLSR analysis was used for obtaining linear regression models that use the variable representing the effort as dependent (namely EFH) and the variable denoting the COSMIC size measure (namely CFP) as independent
CBR (Case-Base Reasoning)

- The idea is to predict the effort of a new project by considering similar projects previously developed.

- Completed projects are characterized in terms of a set of p features and form the case base.

- The new project is also characterized in terms of the same p attributes and it is referred as the target case.

- The similarity between the target case and the other cases in the p-dimensional feature space is measured, and the most similar cases (projects) are used, possibly with adaptations to obtain a prediction for the target case.
• A training set to build the estimation model
• A test set to validate the estimation model

- This is (theoretically) the best option since the test/ validation set is a totally separate dataset from the dataset used to build the effort estimation models
Evaluation Criteria

✓ **goal**: evaluating the accuracy of derived effort estimations
  - MMRE \(\rightarrow\) (mean) where \(\text{MRE} = \frac{|EFH_{\text{real}} - EFH_{\text{pred}}|}{EFH_{\text{real}}}\)
  - MdMRE \(\rightarrow\) (median)
  - Pred(0.25) \(\rightarrow\) Pred (0.25) = \(k / N\)
    - \(k\) is the number of obs having MRE less than 0.25 and \(N\) the number of observation in the validation set
    - is a quantification of the percentage of predictions whose error is less than 25%

✓ **typical expected thresholds** (Conte et al., 1986):
  - MMRE \(\leq 0.25\)
  - Pred(0.25) \(\geq 0.75\)
Agenda

- **Introduction**
  - Early Effort Estimation: motivations
  - Statistical Techniques: OLSR, CBR

- **Sizing and Estimating Web Applications**
  - Using COSMIC
  - Using other Size Measures

- **Data Sets Descriptions & Research Method**
  - Single-company dataset
  - Cross-company dataset
  - Descriptive Statistics
  - Research Method: Techniques & Evaluation Criteria

- **Obtained Results & Data Analysis**
  - Using OLSR
  - Using CBR
  - Comparing the accuracy of obtained estimates

- **Conclusions & Prospects**
• **OLRS assumptions**

  ✓ (single-company dataset) **Satisfied**

  ✓ (cross-company dataset), the variables EFH and CFP were highly skewed and **homoscedasticity** and residual **normality** assumptions were **not satisfied**

  ✓ We transformed the variables by applying the natural log (Ln) on them

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std. Err</th>
<th>t-value</th>
<th>p-value</th>
<th>R²</th>
<th>Std Err</th>
<th>F</th>
<th>Sign F</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Single-company dataset</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coefficient</td>
<td>3.014</td>
<td>0.386</td>
<td>7.801</td>
<td>0.000</td>
<td>0.824</td>
<td>360.099</td>
<td>60.861</td>
<td>0.000</td>
</tr>
<tr>
<td>Intercept</td>
<td>862.281</td>
<td>250.613</td>
<td>3.441</td>
<td>0.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cross-company dataset</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.005</td>
<td>1.621</td>
<td>0.074</td>
<td>0.7889</td>
</tr>
<tr>
<td>Coefficient</td>
<td>0.126</td>
<td>0.4611</td>
<td>2.778</td>
<td>0.015</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>7.608</td>
<td>2.739</td>
<td>0.273</td>
<td>0.789</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

• **Quality of the resulting models**

**Employed indicators:** $R^2$ (close to 1), F-value (high), and Sign F (<0.05) of the model, $t$-value (>1.5) and p-value (<0.05) of the coefficient and intercept

  ✓ (single-company dataset) **Indicators OK!**

  ✓ (cross-company dataset) **Indicators do not hold!**
To obtain “good” effort estimation models

- we cluster the cross-company dataset, by grouping the observations according to a specific parameter: **productivity**
  - We consider two ranges: [0, 1.21] and [1.21, 3], where 1.21 represents the median value of productivity

<table>
<thead>
<tr>
<th>Var</th>
<th>Obs</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>median</th>
<th>STd.dev</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cross-company Set 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EFH</td>
<td>7</td>
<td>6188</td>
<td>46787</td>
<td>19711.4</td>
<td>11165</td>
<td>16515.5</td>
</tr>
<tr>
<td>CFP</td>
<td>7</td>
<td>94</td>
<td>678</td>
<td>286.3</td>
<td>294</td>
<td>202.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cross-company Set 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EFH</td>
<td>8</td>
<td>408</td>
<td>10206</td>
<td>3287.8</td>
<td>2590.5</td>
<td>3169.9</td>
</tr>
<tr>
<td>CFP</td>
<td>8</td>
<td>79</td>
<td>1670</td>
<td>662.8</td>
<td>708.5</td>
<td>505.2</td>
</tr>
</tbody>
</table>

- **Two cross-company datasets**
  - **Set1**: observations having productivity in [0, 1.21]
  - **Set2**: observations having productivity in [1.21, 3]
• The two obtained estimation models:

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std. Err</th>
<th>t-value</th>
<th>p-value</th>
<th>R²</th>
<th>Std Err</th>
<th>F</th>
<th>Sign F</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cross-company Set 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coefficient</td>
<td>0.837</td>
<td>0.366</td>
<td>2.286</td>
<td>0.071</td>
<td>0.511</td>
<td>0.644</td>
<td>5.225</td>
<td>0.071</td>
</tr>
<tr>
<td>Intercept</td>
<td>5.030</td>
<td>2.007</td>
<td>2.507</td>
<td>0.054</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cross-company Set 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coefficient</td>
<td>1.001</td>
<td>0.121</td>
<td>8.288</td>
<td>0.000</td>
<td>0.920</td>
<td>0.313</td>
<td>68.69</td>
<td>0.000</td>
</tr>
<tr>
<td>Intercept</td>
<td>1.521</td>
<td>0.752</td>
<td>2.023</td>
<td>0.089</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

• The model obtained from Set2
  ✓ high $R^2$ value (0.920) and $F$ value (68.69) and low $Sign F$ (<0.001), indicating that the prediction is indeed possible with a high degree of confidence
  ✓ as from the t test, the coefficient is significant at level 0.05, while the intercept at level 0.09 (very close to 0.05)
  ✓ We assessed whether the model could be exploited to estimate the effort of the applications in the single-company dataset (having productivity values in the range [1,3])
• We used the ANGEL tool
  ✓ Euclidean distance as similarity function
  ✓ 1, 2, 3 analogies as number of similar cases to use for estimating
  ✓ mean of k analogy as adaptation technique

• First run
  ✓ Case base: initial cross-company dataset
  ✓ Target case: each observation in the single-company dataset

• Second run
  ✓ Case base: 7 observations in cross-company dataset Set2
  ✓ Target case: each observation in the single-company dataset
**Training set:**
- Cross-company Set2
- Cross-company Set1
- Cross-company dataset

**Test set:**
- single-company observations
- single-company observations
- single-company dataset

**Comparison**

- Estimation Accuracy (MMRE, MdMRE, Pred(0.25))
- Estimation Accuracy (MMRE, MdMRE, Pred(0.25))
- Estimation Accuracy (MMRE, MdMRE, Pred(0.25))

**Data sets & Research Methods**

**Validation**

- OLSR / CBR based model
- OLSR based model
- OLSR / CBR based model
<table>
<thead>
<tr>
<th>Validation</th>
<th>MMRE</th>
<th>MdMRE</th>
<th>Pred(0.25)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimating the effort of single-company observations using the cross-company Set2 with OLSR</td>
<td>0.12</td>
<td>0.05</td>
<td>0.80</td>
</tr>
<tr>
<td>Estimating the effort of single-company observations using the cross-company Set1 with OLSR</td>
<td>10.98</td>
<td>11.04</td>
<td>0.00</td>
</tr>
<tr>
<td>Estimating the effort of single-company observations using the cross-company Set2 with CBR</td>
<td>0.59</td>
<td>0.68</td>
<td>0.07</td>
</tr>
<tr>
<td>Estimating the effort of single-company observations using the cross-company dataset with OLSR</td>
<td>0.84</td>
<td>0.61</td>
<td>0.00</td>
</tr>
<tr>
<td>Estimating the effort of single-company observations using the cross-company dataset with CBR</td>
<td>1.84</td>
<td>0.77</td>
<td>0.00</td>
</tr>
</tbody>
</table>

The use of productivity seems to be useful to classify observations of a cross-company dataset with the aim of deriving accurate estimate of observations contained in a single-company dataset, when the estimation technique employed is the OLSR.
Introduction
- Early Effort Estimation: motivations
- Statistical Techniques: OLSR, CBR

Sizing and Estimating Web Applications
- Using COSMIC
- Using other Size Measures

Data Sets Descriptions & Research Method
- Single-company dataset
- Cross-company dataset
- Descriptive Statistics
- Research Method: Techniques & Evaluation Criteria

Obtained Results & Data Analysis
- Using OLSR
- Using CBR
- Comparing the accuracy of obtained estimates

Conclusions & Prospects
Research question

- Several statistical techniques can be used (e.g. OLSR and CBR) for obtaining more indications on the right values to declare in estimates from the analysis of historical datasets
- ...but what about an organization using external data for its estimates?

Case Study

- Two datasets used, comparing a single and cross-company datasets for estimating web applications
  - COSMIC was the FSMM used as a functional size measure, having a better fit for web environments than other FSMM
  - ISBSG r10 was the external repository from which cross-company data were selected

First Results

- Cross-company datasets could be exploited ... but “guidelines” are needed to select the set of observations to employ

Lessons Learned

- Few project data in our own organizations: plan project data gathering and store them at the project closure
- Knowledge cannot be reduced anyway to the solely data storage

Next Steps

- Replications of the case study in other contexts and employing larger datasets
Thanks for your attention!