Modeling the Relationship between Software Effort and Size Using Deming Regression

Nikolaos Mittas
Makrina Viola Kosti
Basiliki Argyropoulou
Lefteris Angelis

Programming Languages & Software Engineering Lab
Department of Informatics
Aristotle University of Thessaloniki, Greece
{nmittas, mkosti, vargyrop, lef}@csd.auth.gr

PROMISE 2010, September 12-13, Timisoara, Romania
Presentation Outline

- Motivation and Objectives
  - Research questions
- Definitions and Description
- Methodology of Experimentation
  - Dataset
  - Evaluation method
- Results and Conclusions
- Directions of Future Research
The Research Problem in Software Cost Estimation (SCE)

- Accurate prediction of software cost needed
- Plethora of prediction methods
  - Expert judgment to statistical models and machine learning techniques
- The General Research problem
  - Identification of the “best” prediction technique for a certain dataset
- Findings from literature contradictory - no global answer
  - Lack of standardization in SCE methodologies, measurement, reporting techniques, terminology, etc
  - Accuracy depends on dataset
Motivation

- *Regression Analysis* (RA) (especially *Ordinary Least Squares* OLS) → Well-known modeling technique:
  - Various forms of regression for modeling the relationship between effort and size
- Despite the popularity of OLS in SCE → Several restrictions:
  - OLS it is assumed that the values of the independent variable (i.e. *size*) are measured without errors
- The assumption of error-free measurement is not so realistic in SCE:
  - Software size is the result of a counting and estimating process derived from a tool or an expert (*Source Lines of Code* (SLOC) or *Function Points* (FP))
Related Work

- **Literature Regression forms**
  - Miyazaki et al. (1991)
    - Regression based on relative errors
  - Chen & Stromberg (1997)
    - Heteroscedasticity (LMS, Least Median of Squares – LTS, Least Trimmed Squares)
  - Pickard et al. (1999)
    - RR, Robust Regression – LAD, Least Absolute Deviation

- **Inaccuracy of measurement**
  - Miyazaki et al. (1994)
  - Foss et al. (2001)
Proposed Model

- Deming regression → Improvement of the process for modeling the relationship between effort and size
- Deming regression → General class of errors-in-variables models:
  - Appropriate in situations where random errors exist in the measurements of both the independent and the dependent variable
Modeling the SCE Procedure

- SCE is the procedure of predicting the cost of a new project
- Let:
  - $Y \rightarrow$ Real random dependent variable representing the cost of projects
  - $X \rightarrow$ Real random variable representing the size of projects
- Goal to find a regression function
  \[ y_i = f(x_i) + \varepsilon_i \quad (i = 1, \ldots, n) \]
  - $\varepsilon_i \rightarrow$ Real random error
Parametric Estimation Technique

- **Ordinary Least Squares (OLS) Regression**
  \[ y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \]

- **Objective:**
  - Minimization of the overall *Sum of Squared Residuals* (SSR)
    \[ SSR = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \]

- **Final estimation of constant** \( (\beta_0) \) **and slope** \( (\beta_1) \)
  \[ \hat{\beta}_1 = \frac{n \sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} \]
  \[ \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \]

  where
  \[ \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \]
  \[ \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \]
Proposed Model

- Deming Regression:
  - A form of *errors-in-variables* model
  - Takes into account:
    - The error arising from the **dependent** variable *(effort)* (similarly OLS)
    - The error in measurement of the **independent** variable *(size)*
Theoretical Model (1/2)

- **Notation**
  - \((X_i, Y_i) \rightarrow \text{True unknown observations}\)
  - \((x_i, y_i) \rightarrow \text{Erroneously measured observations}\)
  - \((\varepsilon_i, \delta_i) \rightarrow \text{Error terms}\)

- **Constant ratio** \(\lambda\) **of error variances**
  \[
  \lambda = \frac{S^2_{\varepsilon X}}{S^2_{\delta Y}}
  \]

- **Deming Regression Model**
  \[
  x_i = X_i + \varepsilon_i \quad \quad \quad y_i = Y_i + \delta_i
  \]
Theoretical Model (2/2)

- Deming Regression

\[ Y_i = \beta_0 + \beta_1 X_i \]

- Objective:
  - Minimization of the overall weighted Sum of Square Residuals (SSR)

\[
SSR = \sum_{i=1}^{n} \left( \frac{\varepsilon_i^2}{S_{\varepsilon x}^2} + \frac{\delta_i^2}{S_{\delta y}^2} \right) = \\
\sum_{i=1}^{n} \left( (y_i - \beta_0 - \beta_1 X_i)^2 + \lambda (x_i - X_i)^2 \right)
\]
Estimation of Coefficients

- Final estimation of constant ($\beta_0$), slope ($\beta_1$) and $X$

$$
\hat{\beta}_1 = \frac{s_{yy} - \lambda s_{xx} + \sqrt{(s_{yy} - \lambda s_{xx})^2 + 4\lambda s_{xy}^2}}{2s_{xy}}
\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}
$$

$$
X_i = x_i + \frac{\hat{\beta}_1}{\hat{\beta}_1^2 + \lambda}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)
$$

where

$$
\begin{align*}
s_{xx} &= \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \\
s_{yy} &= \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2 \\
s_{xy} &= \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})
\end{align*}
$$
Methodology (Datasets)

- **Desharnais**
  - 77 completed software projects from a Canadian Software house
  - The project size measured with *FPs*

- **COCOMO81**
  - COCOMO81 → Public domain database
  - The project size is measured with *SLOCs*

- **Maxwell**
  - 62 projects from a commercial Finnish bank
  - The project size measured with *FPs*

- **Nasa93**
  - 93 projects from different centers
  - The project size measured with *SLOCs*

For better fitting, size and effort were logarithmically transformed
Methodology (Accuracy measures)

- The predictive accuracy of the cost model is usually based on local measures of error
  - Absolute Error (AE) \[ AE_i = |Y_{Ai} - Y_{Ei}| \]
  - The Magnitude of Relative Error (MRE) \[ MRE_i = \frac{|Y_{Ai} - Y_{Ei}|}{Y_{Ai}} \]

- The global accuracy measures are:
  - Mean and Median Magnitude of Relative Error (MMRE, MdMRE)
  - Mean and Median of Absolute Error (MAE, MdAE)
  - Percentage of projects with MRE\( \leq \)25\% (pred25)
Methodology (Graphical Comparison)

- Graphical analysis through *Regression Error Characteristic* (REC) curves
  
- 2-dimensional plot:
  - x-axis → the error tolerance of a predefined accuracy measure
  - y-axis → the accuracy of a prediction model

- Trade-off between accuracy and tolerance:
  - The accuracy of a model increases as the error tolerance becomes higher
  - $e=0$ → only the predictions that are identical to actual values considered accurate

Accuracy (e) = \[ \frac{\#\text{(projects with error} \leq e)\text{)}}{\#\text{(projects)}} \]
Desharnais Dataset

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>Deming</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>2.993</td>
<td>-2.212</td>
</tr>
<tr>
<td>slope</td>
<td>0.929</td>
<td>1.868</td>
</tr>
</tbody>
</table>

Deming (solid line) vs. OLS (dashed line)
Accuracy Measures (Desharnais)

<table>
<thead>
<tr>
<th>Metric</th>
<th>OLS</th>
<th>Deming</th>
<th>Improvement (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE</td>
<td>2101.13</td>
<td>614.25</td>
<td>70.77%</td>
</tr>
<tr>
<td>MdAE</td>
<td>1107.46</td>
<td>335.59</td>
<td>69.70%</td>
</tr>
<tr>
<td>MMRE (%)</td>
<td>66.88</td>
<td>15.28</td>
<td>77.15%</td>
</tr>
<tr>
<td>MdMRE (%)</td>
<td>34.88</td>
<td>11.30</td>
<td>67.60%</td>
</tr>
<tr>
<td>pred25 (%)</td>
<td>35.06</td>
<td>87.01</td>
<td>148.17%</td>
</tr>
</tbody>
</table>

- Deming outperforms for all accuracy measures
- The improvement ranges from 67.60% (MdMRE) up to 148.17% (pred25)
- The Wilcoxon test statistically signifies the difference for AEs
- Error reduction achieved by Deming
## COCOMO81 Dataset

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>Deming</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>1.204</td>
<td>0.243</td>
</tr>
<tr>
<td>slope</td>
<td>1.106</td>
<td>1.404</td>
</tr>
</tbody>
</table>

Deming (solid line) vs. OLS (dashed line)
Accuracy Measures (COCOMO81)

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>Deming</th>
<th>Improvement (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE</td>
<td>455.37</td>
<td>222.22</td>
<td>51.20%</td>
</tr>
<tr>
<td>MdAE</td>
<td>63.90</td>
<td>22.47</td>
<td>64.84%</td>
</tr>
<tr>
<td>MMRE (%)</td>
<td>137.38</td>
<td>32.99</td>
<td>75.99%</td>
</tr>
<tr>
<td>MdMRE (%)</td>
<td>63.97</td>
<td>26.31</td>
<td>58.87%</td>
</tr>
<tr>
<td>pred25 (%)</td>
<td>19.05</td>
<td>49.21</td>
<td>158.32%</td>
</tr>
</tbody>
</table>

- Deming outperforms for all accuracy measures
- The improvement ranges from 51.20% (MAE) up to 158.32% (pred25)
- The Wilcoxon test statistically signifies the difference for AEs
- AE distribution of OLS → High variability
- REC curve for Deming dominates
Maxwell Dataset

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>Deming</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>3.517</td>
<td>2.088</td>
</tr>
<tr>
<td>slope</td>
<td>0.827</td>
<td>1.065</td>
</tr>
</tbody>
</table>

Deming (solid line) vs. OLS (dashed line)
Accuracy Measures (Maxwell)

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>Deming</th>
<th>Improvement (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE</td>
<td>3766.83</td>
<td>1856.38</td>
<td>50.72%</td>
</tr>
<tr>
<td>MdAE</td>
<td>1997.54</td>
<td>1068.19</td>
<td>46.52%</td>
</tr>
<tr>
<td>MMRE (%)</td>
<td>55.33</td>
<td>25.46</td>
<td>53.99%</td>
</tr>
<tr>
<td>MdMRE (%)</td>
<td>45.22</td>
<td>22.67</td>
<td>49.87%</td>
</tr>
<tr>
<td>Pred25 (%)</td>
<td>20.97</td>
<td>56.45</td>
<td>169.19%</td>
</tr>
</tbody>
</table>

- Deming outperforms for all accuracy measures
- The improvement ranges from 46.52% (MdAE) up to 169.19% (pred25)
- The Wilcoxon test statistically signifies the difference for AEs
- AE distribution of OLS → High variability with a long upper tail
- REC curve for Deming dominates → Solid line climbs rapidly to 1
NASA93 Dataset

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>Deming</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>1.977</td>
<td>1.277</td>
</tr>
<tr>
<td>slope</td>
<td>0.920</td>
<td>1.107</td>
</tr>
</tbody>
</table>

Deming (solid line) vs. OLS (dashed line)
Accuracy Measures (NASA93)

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>Deming</th>
<th>Improvement (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE</td>
<td>346.51</td>
<td>198.94</td>
<td>42.59%</td>
</tr>
<tr>
<td>MdAE</td>
<td>70.34</td>
<td>34.21</td>
<td>51.36%</td>
</tr>
<tr>
<td>MMRE (%)</td>
<td>65.79</td>
<td>26.77</td>
<td>59.31%</td>
</tr>
<tr>
<td>MdMRE (%)</td>
<td>36.08</td>
<td>16.02</td>
<td>55.60%</td>
</tr>
<tr>
<td>pred25 (%)</td>
<td>33.33</td>
<td>64.52</td>
<td>93.58%</td>
</tr>
</tbody>
</table>

- Deming outperforms for all accuracy measures
- The improvement ranges from 42.59% (MdAE) up to 93.58% (pred25)
- The Wilcoxon test statistically signifies the difference for AEs
- AE distribution of OLS → Slightly higher variability
- REC curve for Deming dominates
Conclusions (1/2)

- Study of modeling the relationship between effort and size

- Main idea:
  - OLS is applied under the assumption that the observed values of the variables are measurements which coincide with the true values
  - Not realistic assumption in SCE → Heterogeneous projects with respect to:
    - Nature
    - The way they were measured
Conclusions (2/2)

- **Goal of this paper:**
  - Application of Deming regression
  - Alternative robust technique → Beneficial:
    - Counting process of the size is characterized by uncertainty due to:
      - Subjective decisions of the practitioners
      - Tools

- **Significant improvement compared to OLS:**
  - Several accuracy measures
  - Graphical inspection (REC curves, boxplots)
  - Statistical tests (Wilcoxon matched paired)
Future Work (1/2)

- Method deserves a deeper and thorough study
- Construction of *Prediction Intervals* (PI) → “optimistic” and “pessimistic” guess for the true magnitude of the cost:
  - Researchers suggest that PI → Realistic estimate accounting for both uncertainty and risk
  - Under the assumption of error in measurement → Point estimate is meaningless:
    - Expresses not the response to the true size value, but the response to the measured value
Future Work (2/2)

- Introduction of more explanatory (or independent) variables in the model → Increase the percent of variability of the effort that is explained by the cost function

- Examination of the performance of the comparative models to different situations:
  - Systematic treatment through simulation
  - Errors of the independent variable ranges from a small amount into a high source of variability
Thank You