CHAPTER 6

MODEL ANALYSIS AND VALIDATION

6.1 INTRODUCTION

The criteria that are used for evaluating the software cost estimation models it was defined by B. W. Boehm they are

**Stability and scope** – Whether change in input parameters produce change in the output cost and also whether the user need is catered with an eye on the coverage of it.

**Usability issues** – Whether this model can be used by people easily.

**High definition** - The definition should clearly tell the cost associated and the cost to be excluded.

**Fidelity** – Whether the actual cost is reliable and close to estimates.

**Objectivity** – whether this model eliminates allocating cost variances to poorly calibrated biased factors.

**Constructiveness** – Does this model provides a hygienic estimate with the clear picture about what has to be done and to increase the understandability of the user towards the work done.
**Detail** – Will this easily accommodates the estimation of a software system consisting of a number of subsystems and units. Whether accurate phase and activity breakdown is performed.

**Prospectiveness** – Does the model keep away from the use of information that will not be used till the completion of the project?

**Frugality** – Parsimony means redundancy factors whether eliminated and factors included should contribute to the result.

The ability of the COCOMO II and proposed hybrid estimation models are examined and compared using 8 different performance factors.

### 6.2 **MAGNITUDE OF RELATIVE ERROR (MRE)**

The MRE rate is very important for the enterprise because underestimation and overestimation of effort may harm the enterprise schedule and associated activities. Underestimation may lead the software development team will be under considerable pressure to finish the product quickly or in a hurry and over estimation may lead too many resources will be committed to the project. Though many membership functions were used in literature to represent the cost drives, many of them are not appropriate to clarify the vagueness in the cost drives.

First, calculate the Magnitude of Relative Error (degree of estimating error in an individual estimate) for each data point. This step is a precedent to the next step and is also used to calculate PRED(n). Satisfactory results are indicated by a value of 25 percent or less.

\[
\text{MRE} = \frac{|\text{Predicted} - \text{Actual}|}{\text{Actual}} \quad (6.1)
\]

where,
‘Predicted’ is the predicted effort in person-months using effort estimation model to develop the software.

‘Actual’ is the actual effort in person-months incurred to develop the software.

Figure 6.1 illustrates that 47 projects were over estimated and 49 projects were under estimated. The average MRE rate was found to be 33.65 and hence the result of the effort estimation of COCOMO II is uncertain or because uncertainty of input. By analyzing the 96 projects dataset we have arrived with two set of outputs one in under estimation and another is over estimation.

From the Table 6.1 it is noted that 49 projects were under estimated and 47 projects were over estimated. Out of these two set of results we relaxed the MRE to -5.0 to 4.9 because prediction never comes true, 59 projects lies between -5 to 4.9 and the estimated effort it is acceptable for those 59 projects. And thus the 59 projects’ effort estimation was accepted. Finally the Table concludes that 6 projects were under estimated and 31 projects were over estimated. Considering the 31 project, where the over estimation occurs either in the project data collection or in the cost drives or SF or in the COCOMO II model. Detail analysis of NASA project dataset was carried out for the over estimation problem and finally it is concluded that the problem emerges from the Cost drivers, SF and COCOMO II model. For this reason our research drove us to redefine the effort multiplier.
Table 6.1 MRE analysis of COCOMO II and proposed model

<table>
<thead>
<tr>
<th>Sl. No</th>
<th>MRE Range</th>
<th>COCOMO II</th>
<th>Proposed Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>No. of Projects</td>
<td>Acceptance</td>
</tr>
<tr>
<td>1</td>
<td>&gt;50</td>
<td>1</td>
<td>Under Estimated</td>
</tr>
<tr>
<td>2</td>
<td>30 to 50</td>
<td>1</td>
<td>Under Estimated</td>
</tr>
<tr>
<td>3</td>
<td>10 to 29.9</td>
<td>1</td>
<td>Under Estimated</td>
</tr>
<tr>
<td>4</td>
<td>5 to 9.9</td>
<td>3</td>
<td>Under Estimated</td>
</tr>
<tr>
<td>5</td>
<td>-5 to 4.9</td>
<td>59</td>
<td>Accepted</td>
</tr>
<tr>
<td>6</td>
<td>-5.1 to -10</td>
<td>5</td>
<td>Over Estimated</td>
</tr>
<tr>
<td>7</td>
<td>-10.1 to -40</td>
<td>10</td>
<td>Over Estimated</td>
</tr>
<tr>
<td>8</td>
<td>-40.1 to -100</td>
<td>10</td>
<td>Over Estimated</td>
</tr>
<tr>
<td>9</td>
<td>&lt; -100</td>
<td>6</td>
<td>Over Estimated</td>
</tr>
</tbody>
</table>

From the Table 6.1 it is noted that 60 projects were under estimated and 39 projects were over estimated. Out of these two set of results we have relaxed the MRE to -5.0 to 4.9 because prediction never comes true, 69 projects lies between -5 to 4.9 and the estimated effort it is acceptable for those 69 projects. And thus the 69 projects’ effort estimation was accepted. Finally the Table concludes that 10 projects were under estimated and 17 projects were over estimated. Figure 6.2 illustrates that 39 projects were over estimated and 60 projects were under estimated. The average MRE rate of 96 projects was 5.791.
The mean magnitude of the relative error, or MMRE, is the average percentage of the absolute values of the relative errors over an entire data set. MMRE results are shown in Table 6.2 in the mean % average test. Given $N$ datasets, the MMRE is calculated as follows: for each data set. According to Conte, the MMRE should also have a value of 25 percent or less.

$$MMRE = \frac{100}{N} \sum_{i=1}^{N} \frac{|Predicted_i - Actual_i|}{Actual_i}$$  

(6.2)
Where $N =$ total number of estimates

The average percentage of the absolute value of the relative errors of COCOMO II over the entire 96 project data set is calculated as 0.3365. The average percentage of the absolute value of the relative errors of proposed model over the entire 96 project data set is calculated as 0.0579. The MRE is the basic parameter for calculating the MMRE. The reason for the large difference in MRE is same as the reason for MMRE

6.4 PRED(n)

PRED(n) is used as complementary criterion to count the percentage of estimates that fall within less than of the actual values. The common used value for ‘n’ is 25% and that a good prediction system should offer this accuracy level 75% of the time. A model should also be within 25 percent accuracy, 75 percent of the time. To find this accuracy rate $PRED(n)$, divide the total number of points within a data set that have an $MRE = 0.25$ or less (represented by $k$) by the total number of data points within the data set (represented by $n$). The equation then is: $PRED(n) = k/n$ where $n$ equals 0.25.

In general, PRED(n) reports the average percentage of estimates that were within n percent of the actual values. Given N datasets, then

$$PRED(n) = \frac{100}{N} \sum_{i=1}^{N} \{ 1 \text{ if } MRE_i \leq n/100 \}$$  

For example, $PRED(30) = 50\%$ means that half the estimates are within 30 percent of the actual.
Figure 6.2 Proposed model MRE rate

The Figure 6.3 illustrates that the Pred(n) for both COCOMO II (green series) and the proposed model (purple series). It is observed that the for PRED(30) = 80.21% for COCOMO II and 97.92% for proposed model.
**6.5 ROOT MEAN SQUARE & RELATIVE ROOT MEAN SQUARE (RMS & RRMS)**

RMS (Root Mean Square): Now, calculate the Root Mean Square (model’s ability to accurately forecast the individual actual effort) for each data set. This step is a precedent to the next step only. Again, satisfactory results are indicated by a value of 25 percent or less.

\[
RMS = \left[ \frac{1}{N} \sum (Predicted - Actual)^2 \right]^{0.5}
\]  

(6.4)

RRMS (Relative Root Mean Square): Lastly, calculate the Relative Root Mean Square (model’s ability to accurately forecast the average actual
effort) for each data set. According to Conte, the RRMS should have a value of 25 percent or less.

\[
RRMS = \frac{RMS}{\sqrt{\sum \text{actual}/T}} \tag{6.5}
\]

The RMS value for COCOMO II is 2254.99

The RMS value for the proposed hybrid model is 51.1603

The RRMS value for COCOMO II is 4.43237

The RRMS value for the proposed hybrid model is 0.100559

6.6 ANALYSIS OF SCALE FACTORS

In the COCOMO II model, some of the most important factors contributing to a project's duration and cost are the Scale factors. To set each scale factor to describe the project; these scale factors are determined by the help of the exponent used in the Effort Equation.

This evidence stresses that COCOMO II effort estimation model yields a suspicious output because of uncertainty at the input level. NASA has estimated several projects using COCOMO II. In our research, we have analyzed 96 NASA project dataset for effort estimation. For these 96 projects dataset to yield the actual effort the total value of SF (Scale factor) of COCOMO II must lie between 46.0273 to -1212.16
Figure 6.4 COCOMO II Scale Factor for 96 Projects

i.e. here the enormous value of SF (Scale Factor) may be due to the values of the effort multiplier. But COCOMO II model has defined scale factor range from 1.01 to 31.62.

Figure 6.4 illustrates the values of scale factor of COCOMO II to yield actual effort, 45 projects’ scale factor lied within the range of COCOMO II scale factor, hence it is confirmed that remaining 51 projects effort prediction will not accurate, because uncertainty in the input will lead to uncertainty output. Hence the research drove us to redefine the scale factor. Now considering the 45 projects, whose scale factor lies with in the COCOMO II scale factor range, even though the 45 projects scale factor lies within the range of COCOMO II scale factor, the predicted effort for the 45
projects, were not satisfactory by cross comparing the Figure 6.1 and Figure 6.4.

Hence the question rises that the uncertainty in the input will lead to uncertainty in the output too and it is true for 51 projects, but for 45 projects there is no uncertainty in the input for the given scale factor, hence the problem shifts to the effort multiplier. In COCOMO II, scale factor and the effort multiplier are the two parameters which decide the accuracy of the estimation model; these two parameters must be tuned to get accurate estimation. The above paragraph examines the scale factor and it is concluded that the uncertainty is due the effort multipliers too. There are 17 effort multiplier in COCOMO II, the unsatisfactory effort prediction for the above said 45 projects is due to these effort multipliers, even though its scale factor comes under the range of COCOMO II. For these reason over research drove us to redefine the effort multipliers.

MARE is recurrently used for finding the Mean Absolute Relative Error between the predicted model and the experimental model from which it is been modelled.

\[
MARE(\%) = \frac{1}{n} \sum_{i=1}^{n} MRE \times 100 (6.6)
\]

where MRE is the ratio of the difference between the actual and estimated.

\[
MRE = \frac{Actual - Estimated}{Actual} (6.7)
\]

Variance Absolute Relative Error is defined as follows

\[
VARE(\%) = Var\left[\frac{E - \hat{E}}{E}\right] \times 100 (6.8)
\]
Figure 6.5 Proposed Model Scale Factor for 96 Projects

From the Figure 6.5, it is noted that 1 projects’ SF does not lie between 0.087 to 3.8999 in the proposed model.

6.7 WILCOXON SIGNED-RANK TEST

The next step will be to test the estimates for bias. The Wilcoxon signed-rank test is a simple, nonparametric test that determines level of bias. A nonparametric test may be thought of as a distribution-free test; i.e. no assumptions about the distribution are made. The best results that can be achieved by the model estimates is to show no difference between the number of estimates that over estimated versus those that under estimated. The Wilcoxon signed-rank test is accomplished using the following steps,
1. Divide each validated subset into two groups based on whether the estimated effort was greater (T+) or less (T-) than the actual effort.

2. Sum the absolute value of the differences for the T+ and T- groups. The closer the sums of these values for each group are to each other, the lower the bias.

3. Any significant difference indicates a bias to over or under estimate.

Another performance measure of a model predicting numeric values is the correlation between predicted and actual values. Correlation ranges from +1 to -1 and a correlation of +1 means that there is a perfect positive linear relationship between variables. And can be calculated as follows:

The correlation coefficient for COCOMO II is 0.6952 and the correlation coefficient for the proposed hybrid model is 0.9985.

\[
\begin{align*}
    P &= \frac{\sum_{i}^{T} Predicted_i}{T}, \\
    a &= \frac{\sum_{i}^{T} Actual_i}{T}, \\
    S_p &= \frac{\sum_{i}^{T} (Predicted_i - p)^2}{T-1}, \\
    S_a &= \frac{\sum_{i}^{T} (Actual_i - a)^2}{T-1}, \\
    S_{pa} &= \frac{\sum_{i}^{T} (Predicted_i - p)(Actual_i - a)}{T-1}, \\
    Corr &= S_{pa}/\sqrt{S_p \ast S_a}
\end{align*}
\]
Table 6.2  Overall comparison between the COCOMO II and proposed model

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Parameters</th>
<th>COCOMO II Effort Estimation</th>
<th>Proposed Effort Estimation</th>
<th>Impact Status of Proposed Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No. of Cost drives</td>
<td>17</td>
<td>17</td>
<td>+ve</td>
</tr>
<tr>
<td>2</td>
<td>No. of Scale Factors</td>
<td>5</td>
<td>5</td>
<td>+ve</td>
</tr>
<tr>
<td>3</td>
<td>Scale Factors range</td>
<td>1.01 to 31.62</td>
<td>0.087 to 2.899</td>
<td>+ve</td>
</tr>
<tr>
<td>4</td>
<td>No. of Project lie within the SF range</td>
<td>45</td>
<td>95</td>
<td>+ve</td>
</tr>
<tr>
<td>5</td>
<td>No. of project does not lie within the SF range</td>
<td>51</td>
<td>1</td>
<td>+ve</td>
</tr>
<tr>
<td>6</td>
<td>Maximum MRE (%)</td>
<td>1,073.51</td>
<td>31.46</td>
<td>+ve</td>
</tr>
<tr>
<td>7</td>
<td>MMRE (%)</td>
<td>33.65</td>
<td>5.79</td>
<td>+ve</td>
</tr>
<tr>
<td>8</td>
<td>RMS</td>
<td>2254.99</td>
<td>51.1603</td>
<td>+ve</td>
</tr>
<tr>
<td>9</td>
<td>RRMS</td>
<td>4.43237</td>
<td>0.100559</td>
<td>+ve</td>
</tr>
<tr>
<td>10</td>
<td>Correlation Coefficient</td>
<td>0.6952</td>
<td>0.9985</td>
<td>+ve</td>
</tr>
<tr>
<td>11</td>
<td>No. of Project over estimated</td>
<td>47</td>
<td>39</td>
<td>+ve</td>
</tr>
<tr>
<td>12</td>
<td>Standard deviation for over estimated projects</td>
<td>168.8</td>
<td>10.17</td>
<td>+ve</td>
</tr>
<tr>
<td>13</td>
<td>Variance for over estimated projects</td>
<td>28,494.42</td>
<td>103.4668</td>
<td>+ve</td>
</tr>
<tr>
<td>14</td>
<td>No. of project under estimated</td>
<td>49</td>
<td>60</td>
<td>-ve</td>
</tr>
<tr>
<td>15</td>
<td>Standard deviation for under estimated projects</td>
<td>11.057</td>
<td>5.479</td>
<td>+ve</td>
</tr>
<tr>
<td>16</td>
<td>Variance for under estimated projects</td>
<td>122.1234</td>
<td>30.03</td>
<td>+ve</td>
</tr>
</tbody>
</table>
All these performance measures (correlation, MMRE, and PRED) address subtly different issues. Overall, PRED measures how well an effort model performs, while the MMRE measures poor performance. A single large mistake can skew the MMREs and not affect the PREDs. Sheppard and Schofield comment that MMRE is fairly conservative with a bias against overestimate while PRED (30) will identify those prediction systems that are generally accurate but occasionally wildly inaccurate.

The COCOMO II and the proposed model has been analyzed and the results of the various performance criteria also compared in the Table 6.2. The accuracy of COCOMO II is uncertain because 45 projects’ scale factor is not in the agreement with the model manual. Where as in the proposed model have only one project’s scale factor lies away from the range. Due to the above said scale factor, the rest of the performance measures are uncertain in the COCOMO II. Except for the under estimated result all other are satisfactory with the proposed model. In the proposed model 60 projects were underestimated were as in COCOMO II is 49. But comparing with the results of other measures and the amplitude of under estimation it is accepted.