The Influence of Leadership and Work Culture on Software Cost Effort

Khaled Hamdan  
UGRU, UAE University, Al Ain, UAE  
khamdan@uaeu.ac.ae

Boumediene Belkhouche  
College of Information Technology, UAE University  
Al Ain, UAE  
b.belkhouche@uaeu.ac.ae

Peter Smith  
University of Sunderland, Sunderland  
peter.smith@sunderland.ac.uk

Abstract—Globalization is adding more dimensions to software effort estimation process. The notions of leadership and culture carry with them highly variable assumptions, and thus, must be explicitly modeled. A new model that incorporates leadership and culture is proposed, elaborated and validated. A survey was undertaken to determine the impact of culture and its effect on the software development process in the areas of project team timeliness, collaboration and team work, leadership characteristics, cultural intelligence, motivation and communication. The use of the Bootstrap method for estimating the effort involved in a given project, along with analogies using real historical data, demonstrates the effectiveness of this approach in surmounting difficulties in describing abstract quantitative variables. Our approach is tested on a cluster sample dataset of 41 cases (projects) collected in 2007 from more than 20 organizations. The results show that the inclusion of leadership and culture in the cost estimation model improves the accuracy of software cost estimation.

Index Terms—Effort estimation, Leadership, Team culture, CBR, Bootstrapping, Ontology

I. INTRODUCTION

Software development teams are becoming less homogeneous and more distributed as a result of globalization. In this new setting, team members tend to possess diverse backgrounds, thus affecting the dynamics and quality attributes of the team. These attributes play a significant role in determining the cost and quality of software projects ([8]; [22]). Moreover, the behavior of an organization and its productivity depend highly on the culture and leadership among the members of the organization ([14]; [27]; [12]; [33]; [28]). Consequently, culture and leadership become critical parameters for software cost estimation.

There has been a continuous search for better models and tools to aid project managers in the cost estimation process ([18]; [17]; [13]). Commonly software cost estimates are based on various methods, such as: Algorithmic Estimation Models COCOMO [8]; SLIM [26], and Function Points [4], Expert Judgment [16] and Case-Based Reasoning ([24]; [1]; [31]). Jørgensen and Shepperd [17] identified over 300 papers on software cost estimation. Many attempts have also been made to identify the effect of individual differences in software developers [22]. However, not enough attention is explicitly given to leadership and cultural issues.

Given the multitude of cost estimation approaches, we chose to concentrate our research on analogy methods, such as those proposed by Shepperd and Schofield [32], and specifically on case-based reasoning (CBR). CBR is an approach used to improve effort estimation by understanding and measuring the similarity between cases ([31]; [5]; [23]). CBR tries to predict an outcome by finding similar cases to the current problem ([1]; [30]). The major strategy of CBR is capturing previous experiences into a case database in order to propose solutions to new problems (cases). The database of past projects is used as a reference point in order to combine actual costs of previous projects for the prediction of the costs of a new project with similar attributes. CBR can be applied either at the project level as a whole, or at the sub-system level.

Our research investigates the hypothesis that organizational culture and project leadership are significant factors in determining accurately the cost of software projects development within the Arab Gulf States. In this region, there is a rapid expansion of IT infrastructure and services, and a generalized use of expatriate labor. These states share key similarities, but differ significantly from the rest of the world [2]. Besides culture, a major difference between Arab leadership and leadership in other nations is to be found in Arab authority values. Our ultimate goal was to develop a...
CBR-based cost estimation model that incorporates leadership and culture parameters. Our approach consisted of several stages. In the first stage, a survey of software development projects within government departments in United Arab Emirates (UAE) was undertaken [15]. In stage two, the analysis of the survey highlighted several parameters, and specifically leadership and culture, which have an impact on cost estimation in this area. In stage three, software cost estimation ontology was developed to provide guidance and to reduce ambiguities. In stage four, a cost estimation model was proposed and its implementation was elaborated. The final stage encompassed the evaluation of our cost estimation model.

Our paper is structured as follows. Section 2 provides a review of the role of leadership and culture in software development. We detail the surveys undertaken, and provide a statistical analysis of the data collected from various organizations. Based on this analysis the model parameters are identified. Section 3 presents the new software effort estimation model. This model augments the CBR model by incorporating culture and leadership. The results and evaluation of the model are presented and discussed. Section 4 uses an example to illustrate our method. Section 5 describes the implementation of the model.

II. CULTURE AND LEADERSHIP

The research literature indicates that leadership and culture are strongly related. Culture plays an important role in people's lives in general, and in organizations in particular. It comprises factors such as knowledge, beliefs, values, traits, experiences, language and religion that make up a community, lifestyle and its way of thinking ([29]; [25]). Numerous authors ([14]; [33]; [28]; [29]) emphasize the role of culture on the organization's propensity for learning. Team performance, behavior and attitudes within and outside the organization are affected by culture. Similarly, leadership impacts productivity. Gardner [14] defines leadership as the process of influencing others to achieve a task by providing purpose, direction and motivation. Hence, to achieve consistent success, it is necessary for organizations to promote a positive, inclusive culture and a supportive leadership.

In an attempt to identify attributes characterizing culture and leadership, we interviewed administrators, managers and project leaders and we surveyed the relevant literature. Figure 1 summarizes our synthesis.

Based on these preliminary findings, a survey to assess leadership and culture was carried out. Each leadership characteristic was categorized into four sub-items that were rated 1 to 9 by the respondents. Then for each of the six main characteristics the average of the sub-Items rating was calculated. Project team culture characteristics were similarly assessed. Thirty eight projects were analyzed. A student's t-test and a one-way analysis of variance (ANOVA) were used to determine significant differences in project attributes according to type of project and organization. Associations between attributes were assessed using Pearson's correlation for quantitative data and the Chi-square test of independence for qualitative data. The Kolmogorov-Smirnov test was used to assess the normality of data. Descriptive statistics for leadership and culture variables are shown below (Tables 1 and 2).
In general, the means and the medians of all the leadership and cultural characteristics are quite high indicating their importance. Regarding the correlations between the leadership characteristics, Table 3 shows that all these characteristics correlate highly ($p<0.001$). The cultural characteristics also show strong correlations (see Table 4). The only exception is 'Team Experience' which seems to correlate only with 'Reward Mechanism' and 'Communications'. It is also interesting to see the correlation between leadership and cultural characteristics (see Table 5). 'Team Experience' correlates only with 'Decision-Making' and 'Communication Skills'. Some leadership and cultural characteristics appear to be more important than others. These characteristics were believed by the respondents to be significant attributes in most cases. This is probably due to the fact that these are innate attributes which are part of the individuals’ characters which have been shaped by interaction with others and by life experience in the community.
### TABLE IV. 
CORRELATION BETWEEN CULTURE CHARACTERISTICS

<table>
<thead>
<tr>
<th></th>
<th>Timeliness</th>
<th>Collaboration</th>
<th>Job Stability</th>
<th>Intercultural Intelligence</th>
<th>Reward Mechanism</th>
<th>Communication</th>
<th>Team Experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timeliness Pearson Correlation Sig. (2-tailed)</td>
<td>1</td>
<td>.700**</td>
<td>.791*</td>
<td>.693*</td>
<td>.520*</td>
<td>.598*</td>
<td>.286</td>
</tr>
<tr>
<td>Collaboration Pearson Correlation Sig. (2-tailed)</td>
<td>.700**</td>
<td>1</td>
<td>.694*</td>
<td>.667*</td>
<td>.686*</td>
<td>.668*</td>
<td>.170</td>
</tr>
<tr>
<td>Job Stability Pearson Correlation Sig. (2-tailed)</td>
<td>.791**</td>
<td>.694**</td>
<td>1</td>
<td>.752*</td>
<td>.449*</td>
<td>.476*</td>
<td>.133</td>
</tr>
<tr>
<td>Intercultural Intelligence Pearson Correlation Sig. (2-tailed)</td>
<td>.693**</td>
<td>.667**</td>
<td>.752**</td>
<td>1</td>
<td>.443*</td>
<td>.453*</td>
<td>.110</td>
</tr>
<tr>
<td>Reward Mechanism Pearson Correlation Sig. (2-tailed)</td>
<td>.520**</td>
<td>.686**</td>
<td>.449*</td>
<td>.443*</td>
<td>.449*</td>
<td>.584**</td>
<td>.303</td>
</tr>
<tr>
<td>Communication Pearson Correlation Sig. (2-tailed)</td>
<td>.588**</td>
<td>.666**</td>
<td>.476*</td>
<td>.453*</td>
<td>.584*</td>
<td>.1</td>
<td>.498</td>
</tr>
<tr>
<td>Team Experience Pearson Correlation Sig. (2-tailed)</td>
<td>.286</td>
<td>.170</td>
<td>.133</td>
<td>.119</td>
<td>.303</td>
<td>.487**</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>.073</td>
<td>.295</td>
<td>.413</td>
<td>.498</td>
<td>.057</td>
<td>.002</td>
<td></td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed).

### III. AN AUGMENTED CBR MODEL

The significance of our approach to CBR is the inclusion of culture and leadership in our model (Figure 2).

![Figure 2. The Augmented CBR](image)

The augmented CBR incorporates specific cultural and leadership characteristics, factors, and issues that were identified as influencing cost estimation. The identified parameters have been categorized into seven groups (see Figure 3): organization line of business, application type, organization type, organizational culture, project leadership, project technical environment, and year of project completion.

![Figure 3. Model Parameters](image)
The cultural and leadership characteristics are measured using a nine-point type scale, where 1 means Not Influential at all and 9 means Highly Influential. The remaining parameters are coded as follows. The Organization Line of Business was measured by a seven-point type scale (code 1 to 7): Medical, Governmental Services, Communication, Public Services, Tourism Services, Education and Oil and Gas. The Application Type was measured on a two-point scale (code 1, 2) with 1 being Core systems and 2 being Support systems. The Organizational type was measured on a three-point type scale (code 1 to 3) with 1 being Project Oriented (project manager has the highest power in making decisions), 2 being Matrix (Project Manager has moderate power in making decisions), and 3 being Functional (project manager has lowest level of power in making decisions). The Project Technical Environment parameter is measured according to the number of Core Users (Backend Users), number of Clients, number of Transactions, numbers of Entities, and Technology (Hardware and Software Infrastructure). The Year of Project Completion measures the duration of the project.

A component of our CBR approach is the Estimation by Analogy method (EBA). The key idea behind the EBA is that similar input data vectors have similar output values [5]. A number of nearest neighbors is sought according to a distance metric to determine the output approximation. The estimation of the outputs is calculated by using the average of the outputs of the neighbors (analogies). It is a procedure consisting of three steps. First, the new class for which the project effort is to be estimated is characterized by a set of attributes common to the ones characterizing previous projects in a historical database. Second, one or more similar projects (neighbors or analogies) from the dataset are identified. Similarities and differences between the different projects’ features and the source case that is nearest the target are identified by measuring the distance between cases. Finally, the values of the neighbor projects are used to produce the estimate (usually by computing their mean). A sample data structure for representing the cases is shown in Figure 4.

<table>
<thead>
<tr>
<th>Effort</th>
<th>Attribute 1</th>
<th>Attribute 2</th>
<th>...</th>
<th>Attribute k</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>(E_1)</td>
<td>(X_{11})</td>
<td>(X_{12})</td>
<td>...</td>
</tr>
<tr>
<td>Case 2</td>
<td>(E_2)</td>
<td>(X_{21})</td>
<td>(X_{22})</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Case n</td>
<td>(E_n)</td>
<td>(X_{n1})</td>
<td>(X_{n2})</td>
<td>...</td>
</tr>
<tr>
<td>New Case</td>
<td>Unknown</td>
<td>(Y_1)</td>
<td>(Y_2)</td>
<td>...</td>
</tr>
</tbody>
</table>

Figure 4. The data structure [5]

A. Effort Estimation Procedure

The following 3-stage procedure (Figure 5) is used to estimate the amount of effort for the new project.

1. Extract historical cases that are most similar to the current one according to a selected metric.
2. Estimate the response variable “effort” for the new case based on the extracted cases.
3. Estimate the precision and carry out validation.

   a. Use the Bootstrap method to get an estimate of the standard deviation:
      - Select (with replacement) the bootstrap samples.
      - Determine the bootstrap replicates of the median (or the mean).
      - Compute the standard deviation of the bootstrap replicates of the median (or the mean).

   b. Use the Bootstrap method to get an estimate of the bias and carry out validation.
      - Compute the bootstrap replicates of the bias for the median (or mean) as the difference between each bootstrap replicate of the median (or mean) and the sample median (or mean) of the dataset used.
      - Compute the bootstrap estimate of the bias as the average of the bootstrap replicates of the bias. The mean bias either shows the overestimation (+) or underestimation of the effort (-). A positive value of the mean bias represents overestimation and a negative value represents underestimation.
      - Validate by correcting for the bias according to its sign (+/-).
B. Similar Cases Extraction

Step 1 uses the k-nearest neighbor algorithm to identify similar cases [3]. The unweighted Euclidean distance measure is the most popular and straightforward distance measure that has been previously used with encouraging results in software engineering cost estimation studies ([6]; [32]). When there are mixed (numerical and categorical) data, a dissimilarity measure is also computed. An advantage of this measure is that it takes into account the missing values of the projects as well. In such a case, the distances are usually calculated using only the available attributes.

In the following, it is assumed that the new project is represented by a vector of attributes \(Y_1, Y_2, ..., Y_n\) and every project \(i\) by the vector \(c_{i1}, ..., c_{in}\) as shown in Figure 4. The distance is then computed using the formula shown in Figure 6.

\[
\text{Sim}(c_1, c_2, F) = \frac{1}{\sqrt{\sum_{j \in P} \text{Feature Dissimilarity}(c_{1j}, c_{2j})}}
\]

- \(F = \text{Number of features}\)
- \(c_1, c_2 = \text{Cases}\)

Figure 6. Nearest Neighbour Algorithm [32]

Determining similar projects that operate in a similar organizational culture and ascertaining an acceptable level of similarity among projects with different organizational cultures is challenging. For example, projects A and B may be similar in decision-making, ability to motivate team members, communications, and project manager experience, yet they may be different in organizational culture. In such a case, the difference in understanding culture can be substituted by a buffer time. The magnitude of impact needs calibration after further studies. For example, at 4-6 on the scale of variables, 10% more man-days need to be added to the new project to accommodate individual differences [20]. The weight on the scale is assigned according to the effectiveness of each parameter in the process.

C. Response Variable Estimation

Once similar cases are identified, the unknown effort (of the new case) is estimated by a location statistic (mean, median) of those “neighbhor” cases. The use of the median is motivated by two reasons:

1. The data we are using here exhibit generally skewed distributions (possibly caused by outliers). This in turn gives biased estimates, and the median is known to be affected less than the mean by this phenomenon.
2. The use of the median for inference purposes requires fewer assumptions than that of the mean.

D. Bootstrapping

The Bootstrap method is used to implement the third step in order to adjust the bias of the EBA estimate of the actual effort. This technique is used to estimate the standard error (SE), bias, confidence intervals, and other measures of statistical accuracy [11]. It provides resampling methods that are widely used to estimate parameters and evaluate the biases and the errors of estimation with little or no assumptions about the underlying distribution.

In the Bootstrap method, the original sample with a distribution \(F\) is sufficiently replicated and the expanded sample with a distribution \(\hat{F}\) is used as the new population. The success of this method is partially due to the fact that \(\hat{F}\) is consistent with \(F\). A sample drawn from the population is used to test the estimators. E.g., an
unknown parameter $\theta = t(F)$ is estimated by $\hat{\theta} = s(x)$ on the basis of a random sample $x = (x_1, x_2, \ldots, x_n)$ from the probability distribution $F$. The procedure used by the Bootstrap method to estimate the standard error of $s(x)$ is shown in Figure 6. Given the observed dataset $x = (x_1, x_2, \ldots, x_n)$, a statistic of interest $s(x)$ to compute, a B bootstrap replication of $s$, say $s(x^*_1), s(x^*_2), \ldots, s(x^*_B)$, where $B$, the number of replications, is a some large number, e.g., around 1000.

**Figure 7. Standard Bootstrap**

Here $S_{\text{bootstrap}}(s(x^*))$ is the sample standard deviation of $s(x^*,k)$, $k = 1, 2, \ldots, B$. Generally, the Bootstrap method gives adequate results for B between 25 and 200 [11]. A bootstrap sample $x^* = (x^*_1, x^*_2, \ldots, x^*_B)$ is a random sample of size $n$ drawn from $F$. For the Bootstrap standard error estimate, it is known that $\lim_{B \to \infty} S_{\text{Bootstrap}}(s(x^*)) = SE_F(\hat{\theta}^*)$. The limit is an ideal, though not fully accurate, estimate of the S.E. of $s(x^*)$. In theory, Bootstrap estimates the standard deviation of the sample median. This is normally given by the standard error of the sample medians($s$), i.e., using the distribution of the median which is not obvious. The Bootstrap estimate of the standard error (calculated for the sample median) gives an easy and practical answer.

**IV. MEASURE ACCURACY AND VALIDATION**

The model was tested on a number of governmental development projects in order to determine its accuracy and appropriateness. Results suggest that closer estimates are obtained when cultural and leadership attributes are included in the estimation model. Specifically, the estimation of actual effort improved in 90% of the support system projects and in 50% of the core system projects, when leadership and cultural attributes were added.

We used the jack-knife method to evaluate the predictive accuracy for our approach [21]. This validation method is an effective useful tool for assessing the error of the prediction procedure. Given a set of completed cases, one of the cases (say the ith case) is removed from the dataset and the remaining cases are used as a basis for the estimation of the removed case.

Two measures of local error (Table 6) are calculated:

1. The magnitude of relative error (MRE) [9].
2. The magnitude of relative error to the estimate (MER) [19].

**TABLE VI.
LOCAL ACCURACY MEASURES**

<table>
<thead>
<tr>
<th></th>
<th>MRE</th>
<th>MER</th>
</tr>
</thead>
<tbody>
<tr>
<td>$MRE = \frac{1}{n} \sum_{i=1}^{n} \frac{</td>
<td>\hat{E}_i - E_i</td>
<td>}{E_i}$</td>
</tr>
<tr>
<td>$MER = \frac{1}{n} \sum_{i=1}^{n} \frac{</td>
<td>\hat{E}_i - E_i</td>
<td>}{E_i}$</td>
</tr>
</tbody>
</table>

The local measures are the basis for the estimation of the global predictive accuracy measures MMRE, predmre25, MMER, and predmer25 (Table 7).

**TABLE VII.
GLOBAL ACCURACY MEASURES**

<table>
<thead>
<tr>
<th></th>
<th>MMRE</th>
<th>MMER</th>
</tr>
</thead>
<tbody>
<tr>
<td>$MMRE = \frac{1}{n} \sum_{i=1}^{n} \frac{</td>
<td>\hat{E}_i - E_i</td>
<td>}{E_i}$</td>
</tr>
<tr>
<td>$MMER = \frac{1}{n} \sum_{i=1}^{n} \frac{</td>
<td>\hat{E}_i - E_i</td>
<td>}{E_i}$</td>
</tr>
</tbody>
</table>

If the MMRE (or MMER) is small, then these are a good set of predictions. A usual criterion for accepting a prediction method as good is that it has a $MMRE \leq 0.25$ (similarly for MMER). The opposite is the case for the predmre25 (or predmer25) accuracy measure. A standard criterion for considering a method as acceptable is $predmre25 \geq 0.75$ (similarly for predmer25).

In order to select the appropriate number of analogies the jack-knife technique is applied from one up to ten analogies and the MMRE, MRE, and pred25 accuracy measures are calculated for each of the cases in the whole dataset. It was decided to use one analogy for the predictions, i.e., a number that minimized the MMRE and gave relatively reasonable results for the measures.

The dataset was split according to application type: “Supporting” applications are the systems which support the internal (shared) services in any organization. These applications are not linked directly to the organization mission and vision; rather they enhance the efficiency, effectiveness, and the performance of the supporting services. Those systems share similar features across the government departments.

“Core” applications exist to help to achieve the mission and vision of the organizations and to satisfy their core purpose. The features of these applications are unique. Organizations with a similar line of business could share similar features.

Next, the (core-support) models were intended to measure predictive accuracy (MMRE, Pred25) with and without cultural and leadership characteristics in the split cases. The functionalities of these systems are different and should be treated separately. The analogy showed significant differences between cases for the support systems of the cases including cultural and leadership characteristics which improved the analogy. The core applications improved the analogy by 50 percent when
the two highest effort cases were removed. There are 19 core projects and 17 support projects in which there are no missing values for the dependent variable. In the presence of correlated independent variables, the regression coefficient may not be meaningful. The negative coefficients in equations do not reflect the true effects of independent variables. The fitting accuracy of the model is presented in Table 8 and Table 9. In order to evaluate the predictive accuracy, the jack-knife procedure was used. Then two different MMRE were calculated:

1. The "fitting" MMRE: this is calculated by the regression procedure in SPSS. The "Unstandardized" predicted values are computed for the data that were used to fit the model and these are in fact the predicted logarithm of the efforts. The computed MREs are therefore given by:
\[ \frac{e^{\text{predicted}} - e^{\text{LN(effort)}}}{e^{\text{LN(effort)}}} \]

The jack-knife MRE is:
\[ \text{MRE} = \frac{e^{\text{predicted}} - e^{\text{LN(effort)}}}{e^{\text{LN(effort)}}} \]

The mean and median of all MRE is the predictive MMRE. After calculating both MMRE and MMER, the corresponding pred25 measure for them.

<table>
<thead>
<tr>
<th>TABLE VIII.</th>
<th>ACCURACY MEASURES FOR THE LINEAR REGRESSION MODEL (SUPPORT SYSTEMS)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MMRE</td>
</tr>
<tr>
<td>Regression</td>
<td></td>
</tr>
<tr>
<td>Fitting</td>
<td>20%</td>
</tr>
<tr>
<td>Predictive</td>
<td>56%</td>
</tr>
<tr>
<td>EbA best</td>
<td>129%</td>
</tr>
<tr>
<td>model (n=8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE IX.</th>
<th>ACCURACY MEASURES FOR THE LINEAR REGRESSION MODEL (CORE SYSTEMS)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MMRE</td>
</tr>
<tr>
<td>Regression</td>
<td></td>
</tr>
<tr>
<td>Fitting</td>
<td>24%</td>
</tr>
<tr>
<td>Predictive</td>
<td>42%</td>
</tr>
<tr>
<td>EbA best</td>
<td>102%</td>
</tr>
<tr>
<td>model (n=9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The comparison of the two models shows that the linear regression model outperforms EBA for all support and core systems measures. On the other hand, analysis of the completed projects, including leadership and cultural attributes appears to provide better results. Regression and analogy performed better when cases were split and selected as core and support systems.

Two methods for estimating the actual effort and total cost both for core and support system projects were presented and their accuracy was evaluated. Results suggest that better estimates are obtained when cultural and leadership attributes are included in the estimation model. Specifically, the estimation of actual effort and cost accuracy improved drastically for both support and core systems, when leadership and cultural attributes were added. Total cost may be used as alternative evaluation for software effort estimation due to its importance and significance in predicting the cost model.

The fitting accuracy of the model is presented in Table 10. In order to evaluate the predictive accuracy, the jack-knife procedure was used. After applying linear regression on the project’s leadership characteristics and project team culture attributes separately, it was concluded that a representative model for the dependent variable LNTotalCost could not be built.

© 2013 ACADEMY PUBLISHER
Estimation by analogy (EBA) is another technique for the prediction of a dependent variable. Various neighbors were tried out and the results of the jack-knife procedure are presented in Table 11. As observed, the optimal number of neighbors varies according to the accuracy measure that needs to be optimized. It would appear that '7 neighbors' is a good choice for the construction of an EBA model.

The comparison of the two models shows that the linear regression model outperforms EBA for MMRE, MdMRE, MMER and MdMER, whereas the opposite is true for the remaining measures. On the other hand, the parametric and non-parametric tests do not provide a statistically significant difference between these measures.

Assume y is a new case with actual effort 582. Regarding the evaluation of the predictive accuracy for EBA method, the jack-knife procedure was adopted [21]. First of all, take the absolute value of (actual – estimate) / actual. After applying analogy to estimate the last project (jack-knife y), analogy finds case 1 to be the most similar and reports that 320 is the estimate. However, the true value is 582. So, the relative error for analogy is (320 – 582) / 582. The MER will be calculated based on the procedure (actual – estimate) / estimate. So, the MER is abs((582 – 320) / 320) for the first project (see Table 12).

In order to select the appropriate number of analogies the jack-knife technique was applied from one up to ten analogies and the MMRE, MRE, and pred25 accuracy measures were calculated for each of the cases in the whole dataset. It was decided to use one analogy for the predictions, i.e. a number that minimized the MMRE and gave relatively reasonable results for the measures. The values of the effort for the selected cases were: 320, 105, 138, 324, 600, 750, 1250, 1295, and 1300. It appeared that '9 neighbors' is a reasonable choice for the construction of the EBA model.
The execution of the macro using Minitab gave: \( s(x) = 600 \) and \( se_{\text{boot}}(s(x^*)) = 348 \). How good is this estimate? Compare it to \( Se \theta(x^*) \), which is obtained from the sampling distribution of \( s(x^*) \) bootstrap replications (see Table 13).

<table>
<thead>
<tr>
<th>OrgNo</th>
<th>Project</th>
<th>Org Size</th>
<th>Org LOB</th>
<th>Org Type</th>
<th>Duration</th>
<th>Tools</th>
<th>DBMS</th>
<th>Decision Making</th>
<th>Communications</th>
<th>Team Exp</th>
<th>Both CL</th>
<th>Actual Effort</th>
<th>MRE</th>
<th>MER</th>
</tr>
</thead>
<tbody>
<tr>
<td>08</td>
<td>Telematics system</td>
<td>60</td>
<td>4</td>
<td>2</td>
<td>15</td>
<td>2</td>
<td>7.0</td>
<td>7.8</td>
<td>6.0</td>
<td>1.00</td>
<td>1.00</td>
<td>30.0</td>
<td>45.0%</td>
<td>81.9%</td>
</tr>
<tr>
<td>07</td>
<td>E-Archive</td>
<td>4,000</td>
<td>2</td>
<td>3</td>
<td>11</td>
<td>1</td>
<td>8.0</td>
<td>8.3</td>
<td>6.0</td>
<td>0.50</td>
<td>1.00</td>
<td>320.0</td>
<td>123.4%</td>
<td>55.2%</td>
</tr>
<tr>
<td>32</td>
<td>Well Prognosis</td>
<td>2,296</td>
<td>7</td>
<td>2</td>
<td>7</td>
<td>3</td>
<td>7.6</td>
<td>7.8</td>
<td>6.8</td>
<td>0.48</td>
<td>2.00</td>
<td>600.0</td>
<td>3.1%</td>
<td>3.0%</td>
</tr>
<tr>
<td>04</td>
<td>Fuder (Data Manage)</td>
<td>1,200</td>
<td>7</td>
<td>1</td>
<td>53</td>
<td>5</td>
<td>8.0</td>
<td>8.0</td>
<td>8.0</td>
<td>0.44</td>
<td>1.25</td>
<td>136.0</td>
<td>76.3%</td>
<td>321.7%</td>
</tr>
<tr>
<td>34</td>
<td>Al Ain Municipality</td>
<td>2,000</td>
<td>2</td>
<td>1</td>
<td>12</td>
<td>1</td>
<td>7.3</td>
<td>5.5</td>
<td>6.0</td>
<td>0.42</td>
<td>1.25</td>
<td>114.6%</td>
<td>53.4%</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>Damjan Insurance</td>
<td>2,296</td>
<td>7</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>7.8</td>
<td>8.0</td>
<td>6.8</td>
<td>0.40</td>
<td>324.0</td>
<td>44.3%</td>
<td>78.6%</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>Project Bus. Env.</td>
<td>200</td>
<td>8</td>
<td>1</td>
<td>24</td>
<td>2</td>
<td>7.3</td>
<td>7.8</td>
<td>6.0</td>
<td>0.40</td>
<td>1.25</td>
<td>122.5%</td>
<td>55.1%</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Database Application</td>
<td>16,000</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>7.0</td>
<td>8.0</td>
<td>7.0</td>
<td>0.39</td>
<td>750.0</td>
<td>28.9%</td>
<td>22.4%</td>
<td></td>
</tr>
<tr>
<td>01</td>
<td>Financial (JD Edwar)</td>
<td>1,900</td>
<td>7</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>8.0</td>
<td>9.0</td>
<td>8.0</td>
<td>0.37</td>
<td>105.0</td>
<td>82.0%</td>
<td>454.3%</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>ERP</td>
<td>1,500</td>
<td>2</td>
<td>2</td>
<td>7</td>
<td>1</td>
<td>8.0</td>
<td>8.0</td>
<td>7.5</td>
<td>0.37</td>
<td>1,200.0</td>
<td>106.2%</td>
<td>51.5%</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>Planning and Dev</td>
<td>167</td>
<td>2</td>
<td>1</td>
<td>16</td>
<td>1</td>
<td>7.5</td>
<td>7.5</td>
<td>7.5</td>
<td>6.0</td>
<td>582.0</td>
<td>17.5%</td>
<td>17.5%</td>
<td></td>
</tr>
</tbody>
</table>

The Bootstrap estimate of the median bias was used to estimate the bias of the sample median (which is 600 here, as seen earlier). A Minitab macro (bootstrap bias), was written to display the bias bootstrap replications along with the sample median and its bootstrap estimate of the bias. The execution of the macro bootstrap bias gave \( \text{bias}_{\text{boot}}(s(x^*)) = 18.5 \) which is the bootstrap estimate of the median bias. The Bootstrap estimate of the bias is as shown (see Table 15):

<table>
<thead>
<tr>
<th>x(m)</th>
<th>p(m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>320</td>
<td>0.0014493</td>
</tr>
<tr>
<td>105</td>
<td>0.0289240</td>
</tr>
<tr>
<td>138</td>
<td>0.1144725</td>
</tr>
<tr>
<td>324</td>
<td>0.2206611</td>
</tr>
<tr>
<td>600</td>
<td>0.2689662</td>
</tr>
<tr>
<td>750</td>
<td>0.2206611</td>
</tr>
<tr>
<td>1250</td>
<td>0.1144725</td>
</tr>
<tr>
<td>1295</td>
<td>0.0289240</td>
</tr>
<tr>
<td>1300</td>
<td>0.0014493</td>
</tr>
</tbody>
</table>

The previous distribution gives \( Se \theta(x^*) = 349.5 \) which is very close to the bootstrap estimate obtained earlier. To have more insight about these estimates, histograms were constructed of the 200 replications used in the bootstrap estimate and the histogram of 200 observations generated from the previous distribution of \( s(x^*) \). These turned out to be similar (see Figure 8).
This positive value suggests that the sample median (of 600) overestimates the actual median, and needs correction. A validation operation is then needed and the corrected median is then \( s(x) \) corrected = 600 – \( \text{bias}\_\text{boot}(s(x')) \) = 600 – (18.5). So the estimate (prediction in fact) of the “effort” for the new project is: \( s(x) \) corrected = 581.5 (very close to the actual effort 582 and the median effort 600). This positive value suggests that the sample median (600) overestimates the actual median, and needs correction. The corrected median is then 582 man-days. The above computations show the extent of the simplicity and usefulness of the bootstrap method in estimating the median effort and its standard error.

The bootstrap subsystem model was developed to enable a simple way of predicting the corrected median effort [10]. This validates the project estimation based on the entry of original analogy effort data, bootstrap sample replication size and effort values. It displays the median bootstrap replications along with the sample median and its standard error. The input interface of the system consists of:

- effort data (number of analogy),
- bootstrap sample replication size value,
- effort values.

The system outputs are replicates of data, mean, median and bias (see Figure 9).

A software tool, called SEEOS (software effort estimation ontology system) that supports the application of an analogy-based method was implemented. This tool provides a flexible interface that allows users to experiment with different project characteristic options. The main functions of SEEOS are the following: (1) defining comprehensive attributes for a project; (2) defining attributes, characteristics and measurement protocol; (3) providing the choice of options to be considered such as cultural factors and leadership; (4) determining which attributes are available to provide better accuracy; and (5) generating most similar projects for the required estimate. The structure of the query system is shown in Figure 10.

![Figure 9. Bootstrap user interface](image)

![Figure 10. The structure of the query system](image)

**V. IMPLEMENTATION OF THE SYSTEM**
A. SEEOS Presentation Layer

The main SEEOS presentation layer consists of two panels; the right side panel and the left side panel (see Figure 12). The left side panel shows the project’s entities along with their attributes, descriptions and values. It consists of a tree structure of different selection keys related to the organization and projects. On the other hand, the right panel shows the selected entities to be estimated by the project manager. It consists of a list which will be populated by the selection keys on the left panel by pressing the “Right Arrow” button located between the two panels. In addition, the right panel has two buttons which are “Compare and Show” and “Remove from the List”.

To find the most similar projects, the project manager can select the project’s attributes from a particular leader in the tree structure; he then clicks the “Right Arrow” button to transfer the project attributes to the right panel. The project manager has the option to remove attributes from the list by clicking the “Remove from the List” button in the right panel. Figure 13 shows an example of keys selected from a list.

There are uncertainties in the way various project terms, variables and factors are interpreted. Two projects that may seem similar may indeed be different in a critical way. Moreover, the uncertainty in assessing similarities and differences means that two different analysts could develop significantly different views and effort estimates.

The project manager is able to select similar existing software projects based on well-understood project similarity features. The SEEOS system establishes a set of common project parameters between different projects and provides a common understanding (ontology) of project parameters and their semantics. It accomplishes this by allowing the project manager to give more semantic content to the new project attributes by selecting the attribute’s value and measurement. Figure 14 shows an example of keys surveyed to be completed.

After populating the list with different keys related to a project, the project manager can compare these keys with the ones in the database for different projects by clicking “Compare and Show”. A dialog box will appear showing the comparison of results between the new project and projects in the database. Figure 15 shows an example of results.
The “Duration” column shows the number of days taken to complete a particular project. The “Effort” column shows the number of man-days involved in completing a project; and the last column, the “Match Factor”, shows the calculated similarity value of a project with that created by the user.

B. SEEOS Business Logic Layer

The SEEOS business logic layer consists of seven Java classes. The “Ontology Tree” is the main class for drawing the main window with a tree of different elements and a list box with different command buttons. “DBBean” is a class used for the database management. “InformationEditor” is a class used to add buttons in particular cells. “ProjComparator” is a class used to sort the values of vectors in ascending order. “Projects” is a class used to store all the values against each project in the Database. “Questions” is a class used to store question and answer values. “TitleRenderer” is a class which keeps track of cells of tables. “UtilityFunctions” is also a class consisting of miscellaneous methods providing different functionalities.

C. SEEOS Database System

The SEEOS database system consists of 23 tables (see Figure 16). These tables contain information about different project attributes. The SEEOS online subsystem contains project information from different organizations and inserts them into the database system.

Each organization has a user name and password (Figure 17). Each organization has to input project data such as organization name, region, organization type such as public (non-profit), private (for-profit), and semigovernment. Also, the organization’s line of business such as medical, governmental services, telecommunications, tourism services, public services, education and oil and gas has to be added.
Other essential information to be added includes organization types such as 'project-oriented' where the project manager has the highest power in making decisions, 'matrix' where the project manager has moderate power in making decisions, and 'functional' where the project manager has the lowest level of power in making decisions. Of equal importance is the SEEOS user interface, as well as the information about the developed project such as the project name, application types (core or support), project duration, the estimated and actual project cost, the estimated and actual effort (man-days), entities such as number of transactions, application specific information (source line of code, process, objects or class diagrams, tables and entities, technology used, application architecture, skill sets and accumulated years of experience), the project’s leadership, and project team cultural characteristics defined with scale point values (Figure 18).

The SEEOS system has been tested on a cluster sample dataset of 41 cases (projects) collected from more than 20 organizations. The test demonstrated the utility of the system, and its capability for providing a more comprehensive model of all characteristics, including for the first time, leadership and culture.

VI. Conclusion

In a global and diverse environment, culture and leadership play an important role in that impacts work performance, and consequently, software cost estimation. To determine the extent of this role, we analyzed statistically responses to questionnaires and interviews. The results showed a strong influence of these two factors on cost estimations. We also integrated culture and leadership in the CBR model. The inclusion of leadership and culture in the cost estimation model constitutes an enhancement and refinement. We demonstrated that the inclusion of culture and leadership improves accuracy prediction. The Bootstrap method was used to validate our model. The implementation of the model offers an effective tool to help managers maintain historical data on past projects and estimate costs for new projects. This tool is being tested by the local industry.

REFERENCES


