International Journal of Research in Engineering, Science and Management
Volume-1, Issue-11, November-2018
www.ijresm.com | ISSN (Online): 2581-5792

Software Project Total Cost Estimation Methods: Statistical Approach

Mrityunjaya D. Hatagundi[1], V. Prashant[2]
[1]*PG Student, Department of Telecommunication Engineering, RVCE, Bengaluru, India*
[2]*Assistant Professor, Department of Industrial Engineering and Management, RVCE, Bengaluru, India*

***Abstract*: Software project management has become a crucial field of research due to the increasing role of software in today's world. When a project is insufficiently supported, the quality and speed of the project will suffer. Improving the functions of project management is a main concern in software development organizations. The essence of this paper is to estimate software project cost by Analogy estimation model, incorporating organizational and intercultural factors was developed and evaluated. Analysis was done to show how such added factors can improve the overall accuracy of estimating the cost of a project. Read is expected to have sound knowledge of Project management, Software Project and Different Statistical approaches.**

***Keywords*: Project Management, Cost Estimation, Software Project, Statistical Approach**

1. Introduction

The budget is a plan that identifies the resources, goals and schedule that allows affirm to achieve those goals. Project cost budgeting involves aggregating the estimated costs of individual schedule activities or work package to establish a total cost baseline for measuring project performance. The project scope statements are prepared prior to the summary budget [8]. Suitable budgeting and cost control are essential ingredients of a successful project. On the other hand, the project manager must be careful to keep costs under control, so that the clients will receive the most value for their money [9].

A problem for cost estimation is the total cost allocation of the product. Because all the activities are being done at different places, the actual supply chain is more complex, and this fact has changed the product cost structure. The parametric cost estimation model particularly suited to the earliest stage of design to cost approach. Indirect costs often referred to as \overhead" costs and cannot easily be allocated to one job or projects. The process of recovering overheads is known as \absorption costing" [10]. Clients for large capital projects are very interested in proposed overhead rates, chargeable to projects and required detailed of how the overheads are to be allocated. Fixed costs are said to be fixed when they remain unchanged and must continue to be incurred, even though the workload fluctuates, (e.g. management salaries, and rent insurance. etc.) [10]. An overall cost estimate for a system is derived from global properties, using either algorithmic or non-algorithmic methods. In that sense, the total cost of product must be known at the early design stage, with the maximum of accuracy in order to simplify the trial and error process.

The accuracy of project estimates can have a dramatic impact on profitability. Software development projects are characterized by regularly over running their budgets and rarely meeting deadlines. Effective software estimation is one of the most important software development activities however it is also one of the most di cult. Under estimating a project will lead to under it, under scoping the quality assurance e ort and setting too short a schedule. That in turn can lead to burnout, low quality, loss of credibility, as deadlines are missed an ultimately to inefficient development e ort that takes longer than nominal. Overestimating a project can be almost as bad. Parkinson's Law is that work expands until available time comes into play. Which means that the project will take as long as estimated even if the project was over estimated. An accurate estimate is a critical part of the foundation of efficient software development [11].

The main objective of this paper is to produce a statistical analysis to ascertain how cultural and leadership factors a ect software project cost estimation. This examines the importance that leaders place on cultural characteristics when it comes to project success. To gain an understanding of software measurement (and ultimately an understanding of the representation of the model), it is important to examine software characteristics in order to assess software maturity and to achieve confirmed software e ort estimation improvement. As the objective of this research is to investigate the impact of various attributes on the total cost of project, this paper is also an attempt to build regression models for predicting e ort based on important project attributes, including cultural and leadership factors.

To achieve our objective, the process of collecting information was conducted in three stages, each using a different survey. The first survey, in which eleven government departments were selected as a Pilot Study, was conducted in Abu Dhabi. The survey was then modified and more attributes were added based on the results of the Pilot Study. This posed research questions as to whether and how culture and leadership factors have an impact on the accuracy of software e ort and cost estimation. The survey results indicated that the respondents from all IT departments in the UAE concur with the significance of each of the cultural aspects covered by the survey in carrying out e ort estimation. This shows that community, organizations and team members are intertwined. They are unable to escape their culture or background; they all affect each other [6].

This paper is organized as follows: Section 2 describes the statistical methodology used. Section 3 describes the distribution of actual total cost, explores the correlation of various attributes with the actual e ort. Presents descriptive statistics for leadership and cultural characteristics. Also regression analysis for estimating actual e ort is discussed and a discussion of the statistical results and a conclusion is given in Section 4.

2. Statistical methodology

The original dataset consisted of survey results from 41 software projects. It was compiled after consultation with private and governmental agencies in the UAE [7]. Some of the aforementioned attributes had missing values. These attributes were not considered in the present analysis which took into account attributes with no or even very few missing values. Since the objective of the study is to investigate the impact of various attributes on the actual e ort, the distribution of the e ort needs to be examined. Two projects were incomplete, so their e orts were not recorded. Another project had an extremely large value and was removed from the analysis as an outlier. The remaining 38 projects were analyzed below. It is worth mentioning that data of software attribute values must be analyzed with care, because software measures are not usually normally distributed.

Several statistical techniques were used to analyze the data. Descriptive and graphical methods were used to explore and describe the values of attributes according to the type of project and organization. A student's t-test and a one-way analysis of variance (ANOVA) were used to determine significant differences in project attributes according to the type of project and organization. Associations between attributes were assessed using Pearson's correlation when data were quantitative and the Chi-square test of independence when data were qualitative. Regression analysis was used to develop an equation for explaining the relationship between actual e ort and software project attributes. In addition, the Kolmogorov-Smirnov test was used to assess the normality of data.

The distribution of 'actual e ort' was highly skewed and not normally distributed. Moreover, there was an outlier. For this reason a logarithmic transformation was applied as is quite common in software cost estimation literature [1]-[4].

3. Measuring project budget: Total cost

During the initial phase of any project, an estimation of the total cost of the project based on several factors plays a crucial role in the overall success of the project. For example, when a software project is being developed an accurate estimation of the total cost implies better understanding and realization of all project aspects such as: the man power needed, their availability and their skill set, details tasks to be completed, the project timeframe and most importantly the choice of any partners or outsources and their commitment. Therefore, an educated estimate of the total cost of a software project will most likely lead to more successful results and on time completion.

The total cost is computed as the total cost for the complete project. This involves each skill set and the amount of work on the project for a period of time (in years). The attribute 'man-year' was calculated by the number of skilled sta and the number of years. In this study, the rate was sourced as in [5], along with a study of comparative salaries. The cost was a multiplication of rate per month by 12 months and by man-years. This was repeated with other skills. The total skill cost was calculated by adding all costs. The person-days (e ort) was calculated in days (see Table 1 below). The days in a year are approximately 185 excluding holidays and could vary in some workplaces.

Table 1
Total actual project cost for sample project

Skills	No	years	Man Years	Days Year	Man Days	rate/M	Cost
Analyst	10	4	40			40000	19,200,000
DB Desn	5	3	15			35000	6,300,000
Network	10	2	20			30000	7,200,000
PM	1	4	4			60000	2,880,000
QA	1	1	1			30000	360,000
Total	27		80	185	14800		$35,940,000

A model was built with the total skill cost as the dependent variable. The descriptive statistics of the dependent variable are presented in Table 2.

Table 2
Descriptive statistics for total skill cost

Total Skill Cost

N	Valid	41
	Missing	0
Mean		2052259
Median		1223040
Std. Deviation		2320646
Skewness		2,470
Std. Error of Skewness		,369
Minimum		160000
Maximum		10535958

Table 3
One sample Kolmogorov-Smirnov test for total skill cost

One-Sample Kolmogorov-Smirnov Test

		Total Skill Cost
N		41
Normal Parameters a,b	Mean	2052259
	Std. Deviation	2320646
Most Extreme Differences	Absolute	,256
	Positive	,256
	Negative	-,207
Kolmogorov-Smirnov Z		1,641
Asymp. Sig. (2-tailed)		,009

a. Test distribution is Normal.
b. Calculated from data.

There are 41 projects with no missing values for the dependent variable. In order to test the null hypothesis that the dependent variable comes from the normal distribution, the One-Sample Kolmogorov-Smirnov procedure (see Table 2 and Table 3) and plots were used. There is evidence that the dependent variable does not follow the normality, therefore transformations of the variables are required. Logarithmic transformation was utilized in order to achieve the normality.

The descriptive statistics of the new dependent variable(LN Total Skill Cost) are presented in Table 4. The One-Sample Kolmogorov-Smirnov procedure shows that the new variable follows the normal distribution.

Table 4
Total actual project cost for sample project

LNTotalSkillCost

N	Valid	41
	Missing	0
Mean		14,0973
Median		14,0169
Std. Deviation		,93225
Skewness		,115
Std. Error of Skewness		,369
Minimum		11,98
Maximum		16,17

After applying the one-way ANOVA test for the LNTotalSkillCost and the nominal variables of the dataset, it can be seen that Application Architecture has an impact on the dependent variable (sig<0.056) (Table 5).

Table 5
One way ANOVA for the LNTotalSkillCost and application architecture

	Sum of Squares	df	Mean Square	F	Sig.
Between Grou	4,890	2	2,445	3,110	,056
Within Groups	29,874	38	,786		
Total	34,764	40			

4. Linear regression: Total skill cost

The dependent variable LNTotalSkillCost was used for the construction of a linear regression model. After the entrance and removal of the Project's Leadership characteristics and Project Team Culture independent variables, the following predictors were selected: Project's Leadership characteristics: Interaction and relationships, Decision-Making and Communications. Project Team Culture: Timeliness, Job Stability and Communications The model explains 48.2 percent of the variability of the dependent variable LNTotalSkillCost (sig<0.000), whereas the coefficients are presented in Table 6. In the histogram of the standardized residuals a slight left skewness is highlighted.

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 LNTotalSkillCost could not be built.

Table 6
Coefficients of the linear model (LNTotalSkillCost)

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	14,848	,839		17,698	,000
	Interaction	,532	,190	,683	2,795	,008
	Decision Making	,921	,209	1,237	4,406	,000
	Communication skill	-1,250	,265	-1,451	-4,720	,000
	Time - Respect	-,245	,132	-,383	-1,858	,072
	Job Stability	-,481	,176	-,616	-2,736	,010
	Communications	,435	,122	,613	3,574	,001

a. Dependent Variable: LNTotalSkillCost

5. Conclusion

In this paper, two methods for estimating the actual report 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 e ort and cost accuracy improved drastically for both support and core systems, when leadership and cultural attributes were added. Total skill cost may be used as alternative evaluation for software report estimation due to its importance and significance in predicting the cost model.

References

[1] L. Angelis, I. Stamelos, A Simulation Tool for Efficient Analogy Based Cost Estimation", Empirical Software Engineering, 5, 35-68, 2000.
[2] L.Angelis, I. Stamelos, and M Morisio, Building a Software Cost Estimation Model Based on Categorical Data", Proceedings of the 7th IEEE International Software Metrics Symposium, pp. 4-15, 2001.
[3] M. Shepperd., M. Jorgensen, "A Systematic Review of Software development Cost Estimation Studies", IEEE Vol. 33, NO, 1, Jan 2007.
[4] B. Kitchenham, L. Pickard, S. MacDonell, M. Shepperd, What accuracy statistics really measure.
[5] V. Janulaitis, Information Technology comparative Compensation Study: Salary Survey", Janco Associates, Inc., Park City, UT, pp. 5-33, 2005.
[6] K. Hamdan, F. Abu Sitta, J. Moses, P. Smith, "An Investigation into the Gulf States Government Approaches to Software Development and E ort Estimation", BCS-SQM 2005.
[7] K. Hamdan, B. Belkhouche, P. Smith, The In uence of Culture and Leadership on Cost Estimation. In Proceedings of the international Conferences on Software Process and Product Measurement, pp. 223-232, Munich, Germany, November 2008.
[8] L. Lin, S. Zheng, X. Zhang, X. Jin, and H. Chi, \Performances improvement in radio over ber link through carrier suppression using Stimulated Brillouin Scattering," Optics Express, vol. 18, No. 11, May 24, 2010.
[9] B. Dunbar, Cost Analysis Division. Oce of Program Analysis and Evaluation. 2008. http://www.nasa.gov/ o ces/ pae/home/index.html.
[10] H. Leung, and F. Fan Zhang, "Software Cost Estimation", Department of Computing, the Hong Kong Polytechnic University, 2000.
[11] M. Camargo, B. Rabenasolo, A. Jolly-Desodt, J. Castelain, Application of the parametric cost estimation in the textile supply chain, GEMTEX EA 2461, France , V 3, I 1, 2003.