International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 06 | June 2020 www.irjet.net p-ISSN: 2395-0072

Cost Estimation of Construction Projects Using Regression Analysis

Sahmi Usman[1], Dinu Thomas[2]

[1]PG Student Indira Gandhi Institute of Engineering and Technology, Kothamangalam, Kerala
[2]Assistant Professor Indira Gandhi Institute of Engineering and Technology, Kothamangalam, Kerala

---

Abstract – *Estimating Construction cost and predicting price escalation are major steps for project owners, estimators, and contractors. The construction cost are always subject to fluctuations that trend toward increasing over the long term, which make the pricing process challenging job. Due to lack of data and information in feasible study of the construction projects, therefore the predicating of the construction cost is very difficult. The objective of this study is to identify the critical factors affecting the cost overrun and obtain statistical models using regression analysis. Regression models are obtained using SPSS Software. A formula is driven to predict the final cost based on past records of key construction costs. The main contribution of this study is providing construction stake- holders with a reliable tool for expecting cost of coming projects, especially with the existing Rates of Inflation.*

Key Words: Estimation, Cost overrun, Regression Analysis, Regression models

1. INTRODUCTION

Cost prediction remains a difficult and complex problem, in spite that, the researchers were still studying and trying different approaches and methodologies to solve it. Preparation of a construction cost estimate for any project is a very complex process. Process of construction cost estimation contains many variable factors. Every variable has to be correctly estimated based on proper study, past experience and research to calculate total project cost of construction. Parametric estimation uses the historical data of projects. In this method, the cost of a project is tried to be expressed in terms of different parameters. The parametric cost estimation models are used to express a dependent variable (cost) in terms of independent variables (parameters). By implementing parametric estimation methods, it is possible to produce conceptual (early) cost range estimates. By range estimating, the risk is captured by giving a range of estimations as a function of desired confidence. When regression models are decided to be used, it is always determining the relations between parameters and project costs.

1.1 Objectives

- Identify the major factors affecting the cost of construction.
- To develop a model to predict project cost more exactly and in an effective way using regression analysis
- Illuminate the major and minor error in estimation process.
- Determine the degree of accuracy for the numerical equation.

1.2 Scope

Historical based estimation is an assumption. Assumption will set some tolerance at the final time and cost. These tolerance can be eliminated using regression technique. Cost variation occur in construction industry due to change of material prices, economy, labor productivity and delay in repair and maintenance. So we can apply the regression technique by updating the existing cost rates.

2. METHEDOLOGY

The detailed methodology available in this study is explained below. The methodology includes the identification of major factors affecting the cost of construction by conducting a questionnaire survey. Further a case study is considered and the estimation data is used to develop regression models to predict the cost of construction.

2.1 Data Collection

In this study, 2 types of data are collected. As the initial step, the previous research papers were reviewed to collect primary data for investigating various factors causing cost of construction and thereby identifying the dependent and independent variables. A questionnaire was then drawn up and was divided into two sections. Section A sought to know the general particulars of the respondents while section B was focused on the effects of cost of construction. The respondents were asked to rank the individual effect of cost of construction based on frequency of occurrence according to their own judgment and local working experience. A four point scale is adopted to facilitate ranking exercise and to facilitate the analysis of the responses, the following numerical values were assigned to the respondent's ratings. They were never-1, rarely-2, sometimes-3, very often-4. Questionnaires were distributed in different firms and 30 were returned as response.

Secondary data is collected for applying regression method. From the primary data, the dependent and independent variable are identified and included in the regression models obtained from secondary data. Case studies of 10 residential projects were considered and estimation data are collected.

The cost of materials and labor charges are included when the cost items are being built (for e.g., in shuttering, centering, transporting, leveling, curing, removal of shutters, dewatering, etc).

2.2 Selection of Variables

After conducting the questionnaire survey the collected data was analyzed and the critical factors affecting cost of construction were identified using Relative Importance Index (RII) method in SPSS software. The most common software for data analysis is Statistical Package for the Social Sciences (SPSS). Also, the researcher believed that SPSS is one among the analytics software's are estimate with confidently what will happen next and will be create smarter choices, improve outcomes and solve complex problems.

A four-point Likert scale of 1-4 was adopted to assess the degree of agreement of each cause of impact. Each factor was given a scale of 1 to 4, so that person could easily express the severity impact i.e, 1 be the lowest and 3 be the highest. The scale for impact was categorized into 4 types which are shown in table 1

Table -1: Assessment Grades

Scale of Impact	
1	Never
2	Rarely
3	Sometimes
4	Very Often
5	Always

The scale value for each factor is obtained and ranked based on Relative Importance Index (RII) of the responses are computed for their impact and significance

$$RII = \frac{\Sigma W}{AN}$$

Where ∑W is the total weight given to each factor by the respondents, which ranges from 1 to 5, A is the highest ranking available (i.e, 5 in this case) and N is the total number of respondents that answered the question. Based on the RII values the factors were ranked and identify the most critical factors. A graph showing major affecting factors are shown in Chart 1.

4 critical factors were identified and the independent and dependent variable is selected. In this case Cost is selected as the dependent variable and quantity is selected as the independent variable.

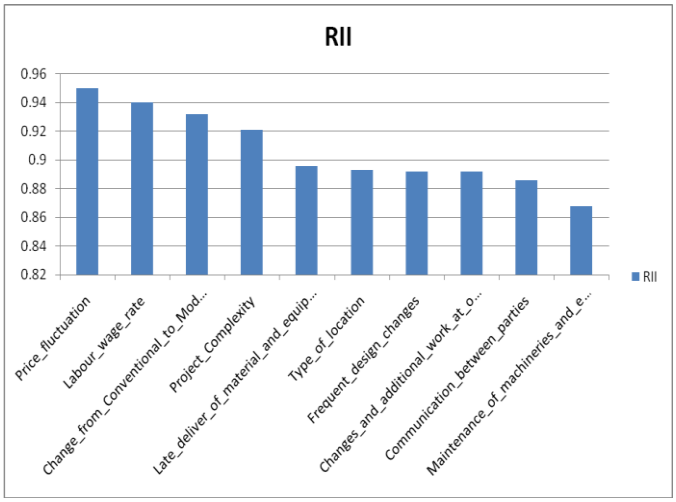

Chart -1: RII value Graph

2.3 Cost Estimation Using Regression Analysis

In statistics, regression analysis is a statistical process for estimating the relationships among variables. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables. A regression analysis is proposed which assumes a cause-and-effect relationship between the variables. In a simple regression, there is one dependent and one independent variable whereas in multiple regressions there is one dependent and at least two independent variables. By using cost data from secondary data collection, regression models were developed. The model developed is used to predict the cost of any amount in the construction projects

3. RESULTS AND DISCUSSION

Both villas and residential building projects are included to fit the model. Each of the project estimates are categorized in to 15 different cost items and shown in Table 2. Separate Excel sheets are created listing out the quantities and their corresponding costs of the various cost items. The sampling technique is done in MS excel sheet representation including from Excavation to MEP works.

Table -2: 15 Different Cost Items

Sl. No.	Item	Unit
1	Earthwork & Site Clearance	m^3
2	Foundation & Plinth	m^3
3	Steel Reinforcement	Tonne

4	Laterite Masonry	m^3
5	Concrete Works	m^3
6	Internal Plastering	m^2
7	External Painting	m^2
8	Tiling	m^2
9	False Ceiling	m^2
10	External Plastering	m^2
11	External Painting	m^2
12	Water Proofing	m^2
13	Joinery	No.s
14	Landscaping	Sq.ft/LS
15	MEP Works	Sq.ft/LS

3.2 Development of models

In this study, linear regression can be used as an optimum forecasting model and very strong mathematical tool that allows the engineers and researchers to find the relationship between dependent variable (cost) and independent variable (quantity). Regression models are used to express a dependent variable in terms of independent variables. The main idea of regression analysis is to fit a curve for the given data while minimizing the sum of squared error and maximizing the coefficient of determination.

The quantity (R), called the linear coefficient of correlation, measures the direction and the strength of a linear relationship between 2 variables. And the coefficient of determination (R^2) represents the percent of the data that is the nearest to the line of best fit. The value of R^2 is determined for each cost items in SPSS Software. As per statistics, if the value of R= 1 then the respective model is highly correlated and R= -1 then the model is highly correlated in reverse fashion i.e. no positive relation can be obtain. In this case, R = 0.99 which is nearly equals to 1, hence the correlation can be found out with very less amount of loss in data. Table 3 shows the model summary of cost item 2 in which correlation coefficient R = 0.921 which represent the cost item is highly correlated with cost of the project.

Table -3: R and R^2 Values of Cost Item 2

Cost Item	R	R^2	Adjusted R^2	Std error of the Estimate
2	0.921	0.99	0.985	1.41368

Chart 2 shows the result of correlation test. The value nearer to 1 means the correlation between dependent variable and independent variable is very strong. The total variation is defined by all the variables are 98% that is the project is feasible from data point of view with only 2% loss of data.

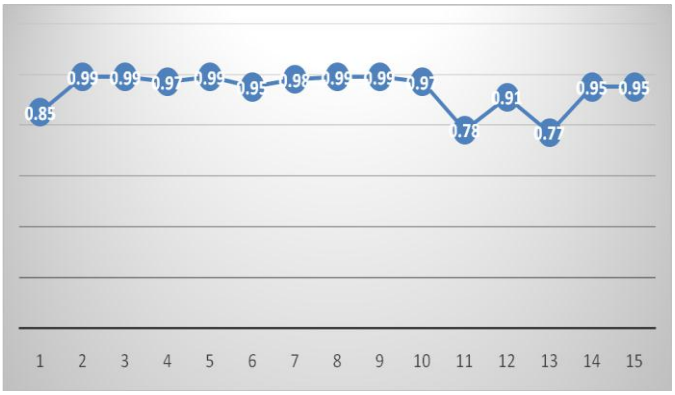

Chart -2: R^2 Values

From the above results, the cost items that are strongly correlated are selected for developing regression models. Linear Regression Analysis generates an equation to explain the mathematical relationship between 2 variables. In statistical science, generally consisting of two variables, the Y variable and X variable. The relationship between two variables (Y and X) will produce a linear regression equation.

$$Y - Y_{Average} = B_{yx} (X - X_{Average})$$

Where B_{yx} is the regression coefficient,

$$Byx = \frac{n \Sigma XV - \Sigma X . \Sigma V}{n \Sigma X2 - (\Sigma X) 2}$$

Regression Coefficient represents that the independent variable have a greater impact on the dependent variable in a correlation analysis, once the variables are measured in numerous units of measurement which are also called slope coefficients. The regression coefficients for each cost items are then determined using the equation of regression coefficient which is given in Table 4. Hence 13 regression models were developed since 2 cost items were rejected based on the result of analysis

Table -4: Regression Coefficient Values

Sl. No.	Item	Regression Coefficient (B_{yx})
1	Earthwork & Site Clearance	181.868
2	Foundation & Plinth	8567.71
3	Steel Reinforcement	18024.511
4	Laterite Masonry	2018.64
5	Concrete Works	17807.124
6	Internal Plastering	253.936
7	Internal Painting	104.455
8	Tiling	480.65
9	False Ceiling	426.648
10	External Plastering	169.290
11	Water Proofing	194.443
12	Landscaping	1772.85
13	MEP Works	497.08

The researchers believed the regression models are the powerful tool for predicting the cost of projects in future depending on the information from past or present projects. The regression models obtained and the Values of regression coefficients can be used to predict the costs.

4. CONCLUSIONS

- Regression Analysis has the ability to estimate the cost of construction projects. It will develop an optimum forecasting model as a comprehensive tool for parametric cost estimation with high degree of accuracy.
- The developed model is useful for engineers and researchers because it provides easy tool to predict the cost of construction projects.
- The result is a model with a realistic method of estimation. This can eliminate the errors usually occurs in parametric estimation methods.
- Using regression analysis method, we establish the statistical relationship between variable parameters. Also able to find relations for more than two independent variables. In Addition, the regression methodologies demonstrated in this study may be used by design and construction professionals to develop their own cost forecasting models using historic construction cost data.
- The accurate prediction of construction cost estimates during the very early stages of project development should enhance overall project success, as it augments early and effective decision making that improve the planning, budgeting, financing and overall management of construction projects. Accurate estimates should improve cost control, project success and client satisfaction.

REFERENCES

[1] Al-Zwainy, F. M. S., Abdulmajeed, M. H. and Aljumaily, H. "Using Linear Regression Technique for Modeling Productivity Construction " Open Journal of Civil Engineering, OJCE, Vol. 3

[2] Smith, S. D, "Earthmoving Productivity Using Linear Regression Techniques," Journal of Construction Engineering and. Management, Vol. 125

[3] Chan, A. P. C., Scott, D., & Chan, A. P, Factors affecting the success of a construction project. Journal of Construction Engineering and Management.

[4] A.,Ayman, A, Hammad , S.M.A. Ali, G. J. Sweis, and A.Bashir, "Prediction Model for Construction Cost and Duration in Jordan", Jordan Journal of Civil Engineering.

[5] E. Elbeltagi, O. Hosny, R.Abdel-Razek, and A. El-Fitory, "Conceptual Cost Estimate of Libyan Highway Projects Using Regression Analysis," International Journal of Engineering Research and Applications, vol.4 .

[6] H.A. Al-Momani, Construction Delay: A Quantitative Analysis. International Journal of Project Management, vol. 18.