Parametric Estimation in Construction using Bootstrap in Regression Models

Vineetha Guruprasad

Abstract— Construction industry is one of the most vast and rapidly growing industries of today. A large part of the income of the society goes into construction of various infrastructures. Hence it is of utmost importance that capital is spent effectively, avoiding the chances for wastage. For this purpose, proper forecasting of cost that maybe incurred would effectively help in assessing the feasibility of a project. Various forecasting methods are now available in the engineering industry. In this paper, the forecasting method used is regression analysis. The construction parameter used here for cost estimation is quantity. 10 various types of construction project details were collected. The various construction phases were grouped into 10 important parameters. The cost and quantity values of these parameters were noted and their relationship was found using regression analysis. Bootstrapping was carried out to increase the validity of models obtained. The models thus produced would help in cost estimation of future projects accurately in the early stages itself. The coefficient of determination (R^2) of the developed models is ranging from 0.5 to 0.99 after bootstrapping. This indicates that the relationship between the dependent and independent variables of the developed models is good.

Index Terms—Bootstrap Approach, Cost, Estimation, Regression Analysis, Quantity.

1 INTRODUCTION

The Construction industry is one of the most vast and rapidly growing industries of today. A large part of the income of the society goes into construction of various infrastructures. Hence it is of utmost importance that capital is spent in an optimized manner. This could be carried out effectively if the capital to be spent can be predicted before the commencement of the project. Proper prediction of cost to be incurred in a project may help contractors to assess its feasibility, decide whether to accept or reject the project, take corrective measures to reduce the expenditure, proper scheduling, etc. There are many methods to forecast the overall cost of projects at various stages. Out of the various methods, most accuracy is obtained in predicting cost from the detailed architectural and structural drawings of the project.

In this thesis, regression analysis is used to attain suitable cost prediction-models. Regression analysis is a statistical approach to find the relation between variables. It helps to determine the nature of relationship between dependent variable and independent variables. It is effective because it is a well-defined mathematical approach. In the present study, the variables used are quantity and cost of various items in the construction project. Here, cost is the dependent variable and quantity is the dependent variable. Since the number of sample projects that can be considered is limited, bootstrapping of regression model is to be employed to obtain the desired accuracy.

Bootstrapping is a general approach to statistical inference based on building a sampling distribution for a statistic by resampling from the data at hand. Once a regression model has been constructed, it is important to confirm the goodness of fit of the model and the statistical significance of the estimated parameters. Model validation is an important step in the modeling process and helps in assessing the reliability of models before they can be used in decision making. Bootstrapping is a model-validation approach. The bootstrap is a computer based method for estimating the standard error of a sample. The bootstrap estimate of standard error requires no theoretical calculations, and is available no matter how mathematically complicated the estimator is. This method will help to increase the accuracy of the results obtained.

2 RESEARCH OBJECTIVE

The cost and quantity per item of 10 building projects were determined by calculating the estimate of the projects. The objective of the paper was to find the relation between quantity and cost per item using regression analysis, apply bootstrap to increase the validity of the results obtained and hence obtain the most apt model for cost-prediction.

3 METHODOLOGY

3.1 Data collection

Estimation requires extensive historical data. The data collected comprises of 10 building construction projects in Kerala. The data were either in the form of prepared estimates or drawings (architectural and structural) from which, detailed estimates were prepared.

3.2 Tabulation

The data collected was grouped and sorted into 10 main and common construction parameters for each project namely:

- Earth excavation
- Foundation
- Superstructure wall
- Reinforced concrete for columns

- Reinforced concrete for beams
- Reinforced concrete for slabs
- Steel reinforcement
- Wooden joinery
- Plastering and painting
- Flooring

These factors were selected based on their rate-analysis and percentage of participation in the overall cost of the project. The various items used for construction in the considered projects were grouped into the above determined factors and the corresponding quantities and cost incurred were calculated and tabulated.

3.3 Analysis

Analysis of tabulated values and selection of appropriate analysis model was the next step. One of the widely used models is the regression estimation model. It is effective due to a well-defined mathematical approach, as well as because of being able to explain the significance of each variable and relationship between 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 (R^2). The models developed in this study are of this type. Linear regression was carried out for each parameter.

3.4 Model Validation

Model validation is an important step in the modeling process and helps in assessing the reliability of models before they can be used in decision making. Analysis models are validated to minimize the model prediction error.

Validating regression model was implemented in this paper by bootstrapping approach**.** The bootstrap is a computer-intensive statistical technique that plays an increasingly important role in modern statistical analysis and applications. By this method, the adjusted R^2 values were noted. If the value was less than 0.5, then quadratic and cubic regression analysis were carried out. The model with higher R^2 value was adopted for that parameter.

4 RESULTS AND DISCUSSIONS

After carrying out linear regression analysis for each parameter, the R^2 values obtained are listed in table 1. The next step after analysis was validation. The results of model validation using bootstrap approach are shown in table 2. From table 2, it is proved that linear regression model does not fit for the construction parameters namely steel reinforcement and joinery work. Hence next step for validation was carried out. These parameters were analysed using quadratic and cubic regression analysis. The analysis results are shown in table 3.

TABLE 1
CONSTRUCTION PARAMETERS AND THEIR R^2 VALUES

Parameter	R^2 value
Earthwork excavation	0.93
Foundation	0.748
Masonry	0.849
RCC for slabs	0.913
RCC for beams	0.944
RCC for columns	0.964
Steel reinforcement	0.249
Joinery	0.251
Plastering and painting	0.950
Flooring	0.813

TABLE 2
BOOTSTRAP RESULTS

Parameter	Initial R^2 value	Adjusted R^2 value	Remarks
Earthwork excavation	0.93	0.921	Linear regression model is feasible
Foundation	0.748	0.717	Linear regression model is feasible
Masonry	0.849	0.830	Linear regression model is feasible
RCC for slabs	0.913	0.902	Linear regression model is feasible
RCC for beams	0.944	0.937	Linear regression model is feasible
RCC for columns	0.964	0.960	Linear regression model is feasible
Steel reinforcement	0.249	0.155	Linear regression model is not feasible
Joinery	0.251	0.157	Linear regression model is not feasible
Plastering and painting	0.950	0.944	Linear regression model is feasible
Flooring	0.813	0.789	Linear regression model is feasible

TABLE 3
ADVANCED REGRESSION RESULTS

Parameter	R^2 value(quadratic)	R^2 value(cubic)	Remarks
Steel reinforcement	0.453	0.574	Cubic regression model is feasible
Joinery	0.923	0.955	Cubic regression model is feasible

Hence the results are summarized in table 4.

TABLE 4
FINAL RESULTS

S.No	Parameter	R^2 value	Regression model
1	Earthwork excavation	0.93	Linear
2	Foundation	0.748	Linear
3	Masonry	0.849	Linear
4	RCC for slabs	0.913	Linear
5	RCC for beams	0.944	Linear
6	RCC for columns	0.964	Linear
7	Steel reinforcement	0.574	Cubic
8	Joinery	0.955	Cubic
9	Plastering and painting	0.950	Linear
10	Flooring	0.813	Linear

Table 5 shows the regression model equations that can be used to predict the cost that could be incurred in future building projects.

TABLE 5
REGRESSION MODEL EQUATIONS

S.No	Parameter	Regression model equation
1	Earthwork excavation	Y = 116.02 X + 20726.96
2	Foundation	Y = 2141.06 X - 198949.25
3	Masonry	Y = 5161.31 X – 253161.17
4	RCC for slabs	Y = 6628.15 X + 32955.11
5	RCC for beams	Y = 6998.38 X -72308.93
6	RCC for columns	Y = 7437.45 X -61401.0
7	Steel reinforcement	$Y = -(3.32 \times 10^{-10}) X^3 + (4.42 \times 10^{-4})X^2 + (1.35 \times 10^{-9})X + (-1.9 \times 10^{-13})$
8	Joinery	$Y = -0.0002X^3 + 0.976X^2 + 0.00075X + (5.48 \times 10^{-7})$
9	Plastering and painting	Y = 202.27 X – 308855.67
10	Flooring	Y = 819.76 X -323668.93

5 Conclusion

In this paper, the costs of the major construction operations were predicted by regression analysis of historical data. For this purpose, data of 10 building construction projects were collected. The quantity and cost of each parameter were calculated after a systematic sorting. Regression analysis was carried out on each construction parameter. The results were validated using Bootstrap approach to find the most apt regression model for the parameter.

The regression model equations eventually formed may help in the forecasting of the costs that may be incurred in future construction projects upto a limit. The forecasted values may only give an approximate knowledge of the cost that can be incurred since it may vary based on the market rates of material and labour, location of site, availability and variability of resources, etc. Even though the models developed are affected by these limitations, they can help contractors effectively to forecast the feasibility of projects and schedule the operations based on detailed study of the models. It may hence, effectively help contractors in decision making, planning and scheduling of future projects.

Acknowledgment

I wish to thank Ms Minu Anna Johny for guiding me throughout the work of this paper.

References

[1] Dr. Hamadu Dallah, "A Bootstrap Approach to Robust Regression", *International Journal of Applied Science and Technology Vol. 2 No. 9;* November 2012

[2] A.I Oredein , T.O Olatayo , A.C Loyinmi, "On Validating Regression Models with Bootstraps and Data Splitting Techniques", *Global Journal of Science Frontier Research Volume 11 Issue 6 Version 1.0;* September 2011

[3] Ibrahim Mahamid, "Conceptual Cost Estimate of Road Construction Projects in Saudi Arabia", *Jordan Journal of Civil Engineering, Volume 7, No. 3*, 2013

[4] Jiehan Zhu, "The Analysis of Bootstrap Method in Linear Regression Effect", *Journal of Mathematics Research Vol. 2, No. 4;* November 2010

[5] Snee, R.O, (1977), Validation of Regression Models, Methods and Examples, Techniques 19. 415 – 428

[6] Deflandre, D., Kleijnen, J.P.C., "Statistical analysis of random simulations: Bootstrap tutorial". Simulation News Europe (38/39), 29-34, 2003.

[7] Sony Sulaksono, "Performance of Bootstrap-Estimated Discrete Choice Models: A Case Study of Bangkok Transit System (Bts)", *Journal of the Eastern Asia Society for Transportation Studies, Vol. 6, pp. 1766 - 1774*, 2005.

[8] Alan O Sykes, "An Introduction to Regression Analysis", 2003.

[9] Huseyin Karanci, "A Comparative Study Of Regression Analysis, Neural Networks And Case – Based Reasoning For Early Range Cost Estimation Of Mass Housing Projects", September 2010.

[10] John Fox & Sanford Weisberg, "Bootstrapping Regression Models in R", June 2012.