International Journal On Engineering Technology and Sciences – IJETS™
ISSN (P): 2349-3968, ISSN (O): 2349-3976
Volume 2 - Issue 5, May 2015

MODELING THE PARAMETRIC CONSTRUCTION PROJECT COST ESTIMATE USING BOOTSTRAP AND REGRESSION TECHNIQUE

[1]Ms. Saira Varghese, [2]Dr. S.Kanchana
[1], [2]Civil Engineering, RVS Technical Campus-Coimbatore/Anna University-Chennai/India

ABSTRACT

This paper presents a parametric estimation approach for the prediction of future construction cost. In the proposed method, a combined regression analysis and bootstrap re-sampling technique is used to develop the range estimate for construction costs. In parametric estimating, a model including the important parameters is developed to predict construction costs, using data of previous projects and the regression analysis is used to determine an adequate cost model. The regression models can be generated using software's like SPSS (Statistical Package for Social Sciences) and MS Excel. This study is of relevance to practitioners and researchers for conceptual estimation of construction costs.

Keywords: *Coefficient of Determination, Construction Project, Cost Significant Model, Estimation Cost, Parametric Estimation Regression Analysis.*

1. INTRODUCTION

Every project has a certain objective (scope), has defined start and end dates (schedule) and has funding limits (budget). This defines the border of a construction project. Since budget and schedule are the main project constraints to be worked with, any estimation concerned with cost and duration are very helpful in the early stages of construction project management process. Construction cost estimating is essential for all of the stakeholders of a construction project, for an ordinary project, namely they are owner, designer, contractor and subcontractor.
Construction cost estimation can be defined as "an effort to forecast the actual cost". Cost estimations can be done in any stage of the project (Feasibility Stage, Conceptual Stage, Engineering Stage, Procurement Stage, Construction & Turnover).

Conceptual (early) cost estimation is performed in Conceptual Stage before detailed design is completed. In Conceptual Stage, the preliminary design of the project has been finished. Preliminary drawings and specifications are the only sources that can be used in conceptual cost estimation. For an accurate estimate, detailed scope definition is essential. At the early stages of a construction project the design information and scope definitions are very limited, hence achieving high accuracy is very difficult. As the project proceeds from Feasibility Stage to Turnover Stage, the accuracy of cost estimating increases due to the finalized drawings and specifications.
Conceptual cost estimating methods are: Unit Cost Method, Factor Method, Probabilistic Modelling & Simulation and Parametric Estimation

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 probabilistic modeling & simulation and 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, there is always the problem of determining the class of relations between parameters and project costs. It is hard to find the accurate relation between dependent (cost) and independent variables (parameters) when there are multiple cost components. Regression model is a parsimonious model. A parsimonious model can be defined as: "a model that fits the data adequately without using any unnecessary parameters".

Since conceptual cost estimations of projects are employed in the early stages, by implementing point estimates, it is not possible to take into account the uncertainty in the estimations. The variability included in the estimations should be emphasized by providing range estimates. In this context, the main purpose of this study is to develop a method for early range estimations of costs by using regression analysis and bootstrap technique. To implement this study, data of a building project was taken with detailed cost estimation and applied to the analysis as the sample set.

2. BOOTSTRAP TECHNIQUE

2.1 History

The bootstrap method introduced in Efron (1979) is a very general re-sampling procedure for estimating the distributions of statistics based on independent observations. The bootstrap method is shown to be successful in many situations, which is being accepted as an alternative to the asymptotic methods such as the traditional normal approximation.

2.2 Approach

Bootstrap is a viable alternative to asymptotic standard errors and confidence intervals. In statistics, bootstrapping is a method for assigning measures of accuracy to sample estimates. This technique allows estimation of the sampling distribution of almost any statistic using only very simple methods. Bootstrap is a re-sampling method in which *r* new samples, each the same size as the observed data, are randomly drawn with replacement from the observed data. The purpose of bootstrap is to mimic the process of sampling observations from the population by re-sampling data from the observed sample. Bootstrap methods are more flexible than classical statistical methods as they require fewer assumptions.

3. REGRESSION TECHNIQUE

In regression problems, *case resampling* refers to the simple scheme of resampling individual cases - often rows of a data set. For regression problems, so long as the data set is fairly large, this simple scheme is often acceptable.

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. A linear relationship between the dependent and independent variables is assumed. An error term U is added in the regression model for capturing the effect of the omitted variables. The estimation of the regression model is carried out by the ordinary least squares (OLS) method. The OLS method aims at minimizing the error sum of the squares while estimating the regression model.

A test is conducted for testing the significance of the individual regression coefficients. The overall fit of the regression is given by R^2 that is called the coefficient of determination and is a measure of the explanatory power of the model. The value of R^2 lies between 0 and 1 (both values inclusive). The closer the value of R^2 to one, the better is the goodness of fit. The significance of R^2 is carried out by using the F statistic. The use of regression in estimating the point and interval prediction is shown. Also is demonstrated the computation of elasticity and its use in decision making.

4. PARAMETRIC BOOTSTRAP AND REGRESSION TECHNIQUE

4.1 Parametric bootstrap

In this case a parametric model is fitted to the data and samples of random numbers are drawn from this fitted model. Then the quantity, or estimate, of interest is calculated from these data. This sampling process is repeated many times as for other bootstrap methods. The use of a parametric model at the sampling stage of the bootstrap methodology leads to procedures which are different from those obtained by applying basic statistical theory to inference for the same model.

4.2 Parametric Methods or Parameters

The cost estimating model used in design phase must be in coherence with the definition report of the project. It is then better to consider the concept of product architecture. We have not known yet how these products will be produced, but we can access a number of physical characteristics or parameters like the mass, the volume, the energy absorbed or the number of input-output. This is obtained for information purposes at the Beginning of development phase.

The parametric estimate has been constructed specifically to create from these parameters the costs to estimate. We must go from high techniques characterizing the pro- duct and owned by the design engineer to economic data necessary for management.

Two main types of parametric methods are to be high-lighted; they differentiated both by their design and usage.

a) The cost list;
b) The statistical models or cost estimates formula.

This method involves more complex calculations than the analogy method. It only works on family products (or projects) of identical structure differing only in size. The calculations are automated on a spreadsheet such as Excel. The use of the application of parametric assessment can be within reach of an unskilled operator .On the contrary, the development of cost estimation formula is complex and requires an expert experienced in statistical calculation.

This method is very practical to conduct appropriate testing in an activity cost analysis. It is appropriate to be used at the time of the unavailability of data and information to make the cost estimation process more accurate. One of parameters method that frequently used to indicate the relationship between costs and the variable is a linear curve.

The formula that shows the cost relationship with independent variables, namely:

$$\mathbf{Y = a\ x}$$

Where, Y = Project cost (dependent variable)
a = Parameter value
x = Variable (independent variable)

Project Cost:- It is the result of a project estimate that is used to determine the relationship between the costs of the various parts of the project.

Parameter Value:- Parameter value is the regression coefficient (can use SPSS). Based on the number of parameters that vary on each building project, then in future studies, we can include the other parameters that are assumed to represent all the parameters that exist in every building project.

These parameters include:
a1 = Parameter value for reinforced concrete work
a2 = Parameter value for roof structure and roof covering work
a3 = Parameter value for timber structure and ceilings work
a4 = Parameter value for sills/frames, doors, windows, and partitions work
a5 = Parameter value for floor work and coating work

Variable:- 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.

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 (R^2).

5. COST SIGNIFICANT MODEL

Viewed from the side of construction, this is the necessary budget plan to complete a project and it is a cost that has been set, based on the results obtained. Plan of budgeting in the construction process is generally intended to estimate the amount of financing a project, not a real cost (actual cost) to be spent. Cost model can be used to organize and distribute the cost estimation or estimated costs into functional areas that can be defined and calculated quantities. The most common ways to find out the cost of a building is a parameter method. This is most often used for project feasibility studies.

The cost model is the latest technique used in cost estimating on a project being proposed. A mathematical model or formula shows that best describe the data collected in the form of cost or price. Another technique often used is single linear regression analysis, which is representative of cost models, despite the fact shows that a more appropriate technique is simulation.

6. SPSS SOFTWARE

The abbreviation of SPSS is "Statistical Package for Social Sciences". SPSS is a Windows based program that can be used to perform data entry and analysis and to create tables and graphs. SPSS is capable of handling large amounts of data and can perform all of the analyses covered in the text and much more. SPSS is commonly used in the Social Sciences and in the business world, so familiarity with this program should serve you well in the future. SPSS is updated often.

7. MICROSOFT EXCEL

Excel is the spreadsheet program created by Microsoft. The instructions given here are specific. The information given in this quick tutorial is meant to give a working knowledge of how to use Excel. There are usually several different ways to perform the same function in Excel and several functions can be performed.

The validation part of the project thesis is done by using Microsoft Excel. The single linear regression equations are created in the spreadsheet and the models are tested in the excel sheet.

8. CONCLUSION

Parametric and probabilistic estimating techniques were integrated using regression analysis and bootstrap methods. The bootstrap method doesn't require any assumptions regarding the probability distribution of the model errors, or the distributions of the cost items, and can be implemented by using a spreadsheet program. Integration of parametric and probabilistic estimating techniques will hopefully lead to more realistic cost expectation during early project stages.

The implementation of bootstrap method for conceptual cost estimating offers several advantages over the classical statistical techniques. The non-parametric bootstrap avoids restrictive and sometimes dangerous assumptions about the form of underlying populations. In addition to cost estimating, potential implementation areas of bootstrap in construction management include risk analysis, probabilistic scheduling, and simulation of construction operations and modelling and hypothesis testing applications.

REFERENCES

[1] Abdelhak Challal and Mohammed Tkiouat (2012) "The Design of Cost Estimating Model of Construction Project: Application and Simulation" *Open Journal of Accounting* July pp. 15-26.
[2] Rifat Sonmez, (2004). "Conceptual cost estimation of building projects with regression analysis and neural networks", *Can. J. Civ. Eng.*, 31(4), 677-683.
[3] Curran M.W. (1989). "Range estimating: Contigencies with confidence." *Transactions of Annual Meeting of Assoc. For Advancement of Cost Engineering (AACE),* B.7.1-B.7.4.
[4] Davidson A.C and Hinkley D.V.(1997). "Bootstrap methods and their applications", *Cambridge University Press,* Cambridge, U.K.

[5] Oberlender, G.D., and Trost, S. M. (2001). "Predicting accuracy of early cost estimates based on estimate quality." *J. Constr. Eng. Manage.*, 127(3), 173–182.
[6] Diekmann J.E. (1983). "Probabilistic estimating: Mathematics and applications" *J. Constr. Eng. Manage.*,
109(3), 297-308.
[7] Ajanlekoko J.O (1987), "Controlling Cost in Construction Industry", *Lagos QS Digest*, Vol.1, No.1, pp.8-12.
[8] Efron.B and Tibshirani. R (1993), *An introduction to bootstrap*, Chapman and Hall, New York.
[9] Sonmez R,(2008). "Parametric Range Estimating of Building Costs Using Regression Models and Bootstrap",*J. Constr. Eng. Manage*, Vol. 134, No. 12. ©ASCE.
[10] Hegazy, T., and Ayed, A. (1998). "Neural network model for parametric cost estimation of highway projects."*J. Constr. Eng. Manage.*, 124(3), 210–218.
[11] Johnson R.W. (2011). "An introduction to bootstrap", *Teaching Statistics*, 23(2), 49-54.
[12] Karshenas, S.(1984) "Predesign cost estimating method for multistory buildings." *J. Constr. Eng. Manage.*,
110(1), 79–86.
[13] Touran, A., and Wiser, E. (1992). "Monte Carlo technique with correlated random variables." *J. Constr. Eng.Manage.*, 118(2), 258–272.
[14] Wang, W. C. _2002_. "SIM-UTILITY: Model for project ceiling price determination." *J. Constr. Eng. Manage.*, 128(1), 76–84.
[15] Lowe, D. J., Emsley, M. W., and Harding, A. (2006). "Predicting construction cost using multiple regressiontechniques." *J. Constr. Eng. Manage.*, 132(7), 750–758.
[16] Scott Armstrong. J (2012), "Illusions in Regression Analysis", *International Journal of Forecasting*.
[17] Mohammed S. Al-Ansari, "Conceptual Cost Estimates for Buildings in Qatar", International Journal of CivilEngineering & Technology (IJCIET), Volume 4, Issue 4, 2013, pp. 284 - 288, ISSN Print: 0976 – 6308, ISSN Online: 0976 – 6316.