A Review on Construction Cost Forecasting Techniques

Shabniya.V¹, Dilruba .K. M²

¹.MTech Scholar, Department of Civil Engineering, MES College of Engineering, Kuttippuram, Kerala, India
².Assistant professor, Department of Civil Engineering, MES College of Engineering, Kuttippuram, Kerala, India

Abstract - Construction industry is considered an important sector for the development all over the World. Cost estimation process is a vital component of the construction industry which adds up to the most significant aspects for the appropriate functioning of any construction company. Construction cost estimation of building construction projects is an essential task in the management of these projects. The quality of construction management depends heavily on their accurate cost estimation. There are mainly two classifications for the cost estimation methods namely Traditional and Non-Traditional. This paper deals with the study of various cost estimation methods, especially the Non- Traditional methods in predicting construction cost of building projects.

Key Words: Artificial Neural Networks, Cost estimation, Case Base Reasoning, Fuzzy Inference system, Monte-Carlo Simulation, Non- traditional methods, Quantity Rate Analysis, Reference class forecasting, Regression Analysis, Support Vector machine.

1. INTRODUCTION

A cost estimate is actually the approximation of cost at the early stage of a particular project or work. Accurate cost estimation a construction project is key factor in a project’s success. But it is hard to estimate construction costs at the planning stage rapidly and precisely, when drawings, documentation etc.. are still incomplete. As such, various techniques have been applied to accurately estimate construction costs at an early stage, when project information is limited. While the various techniques have their problems and consequences, there has been little effort made to determine the best technique in terms of cost estimating performance.

Several estimation methods are used in construction practice and the appropriateness of any particular method is usually reliant on the purpose it is used for, the amount of information available at the time of estimation, and the party using it. Despite the reliance of clients and contractors on available cost estimation and forecasting methods, the actual final costs of construction projects still considerably deviate from their original estimates. Savas Bayram and Saad Al-Jibouri [24][2016]. Afshin Firoz et. al [2] (2016) intended to propose a generic copula-based Monte Carlo simulation method for prediction of construction projects’ total costs with dependent cost items. Found that different dependence structures can lead to different probability distributions of total cost. Also finds that the existing goodness of fit tests can be employed in choosing the best performing copula. The paper concluded that the copula-based Monte Carlo simulation method can predict total cost of construction projects with reasonable accuracy. Onur Dursun and Christian Stoy [20][2016] adopted a novel approach from the domain of forecasting. The multistep ahead (MSA) approach relies on the idea of using several cascaded estimations to predict future values. Accordingly, building element quantities were estimated as the first step. In the second step, estimated quantities were combined with the existing set of inputs to achieve a higher accuracy in construction cost prediction. In order to test the hypotheses of interest, 657 building projects from Germany were analyzed using linear regression and artificial neural network methods. Conclusive evidence suggests that the MSA approach significantly outperforms the prediction accuracy of the conventional practice. Wismulsvara et. al [29][2015] conducted a study on Hybrid Artificial Intelligence on Schematic Design Stage: RANFIS and CBR-GA and compared the performance of cost
estimation models of two different hybrid artificial intelligence approaches: regression analysis-adaptive neuro fuzzy inference system (RANFIS) and case based reasoning-genetic algorithm (CBR-GA) techniques. Models were developed based on the same 50 low-cost apartment project datasets in Indonesia. Tested on another five testing data, the models were proven to perform very well in term of accuracy. A CBR-GA model was found to be the best performer but suffered from disadvantage of needing 15 cost drivers if compared to only 4 cost drivers required by RANFIS for on-par performance. Zhongfu Qin et. al [30] (2015), conducted a Research on Forecasting the Cost of Residential Construction Based on PCA and LS-SVM. He forecasted the costs of a residential construction rapidly and accurately in the initial stage of construction and lack of relevant information. Based on the strengths and weaknesses of previous studies about it, a new model to forecast the costs of a residential construction which is based on Principal Component Analysis (PCA) and Least Squares Support Vector Machine (LS-SVM) is proposed. Finally, selected 5 projects in conjunction with the new model for simulation analysis, the relative error are controlled with $\pm 7\%$. Robert F. Bordle, [21] (2014), studied on “Reference Class Forecasting: Resolving it’s Challenge to Statistical Modeling” and he used a Bayesian approach for combining the reference class forecast and the model-based forecast. He used this approach to estimate healthcare costs under a Voluntary Employee Benefit Association (VEBA). Gwang-Hee Kim et. al [10] (2013) compared the accuracy of three estimating techniques (regression analysis (RA), neural network (NN), and support vector machine techniques (SVM)) by performing estimations of construction costs. By comparing the accuracy of these techniques using historical cost data, it was found that NN model showed more accurate estimation results than the RA and SVM models. Consequently, it is determined that NN model is most suitable for estimating the cost of school building projects.

3. PROBLEMS ENCOUNTERED

Proficient application of the cost estimation process throughout the construction project life cycle is vital to obtaining the goal of a quality cost estimate. Although extremely important and often ignored, project cost estimating has many barriers to conquer throughout the process. Some common barriers include:

- Cost Overruns: Cost overrun is defined as the difference between the final, completed cost of a project and its initial cost estimate. If this continues for a period of time, it may affect the construction projects, resulting in budgetary disorder.
- Schedule Delays: Delay of Schedule can result from many factors and can happen at any time during project construction. Sometimes it could not be removed from the construction process and must be deal with and cause the variation in budget.
- Changes in scope: Changes in scope means accompaniments to the initial plan or idea for a project not firstly discussed as part of the original plan or concept. The possibility that additional items added at any time throughout the project’s life cycle causes severe distress.
- Contingencies: Contingencies are defined as money that has been added in addition to the final cost estimate as a provision for unforeseen instances, such as weather delays and/or changes in scope etc. A certain contingency fund should be included in the initial project cost estimate.
- Inflation: Inflation is an increase of expenditure levels resulting from a substantial and extended rise in prices and other costs through time without changes in project scope. It may also cause financial problems.

4. COST ESTIMATION TECHNIQUES

The cost estimating techniques used could mainly be classified into two categories - Traditional methods and Non-traditional methods. Traditional methods are those methods which are used commonly for the cost estimation of building projects by the estimators. The main traditional method used for building cost estimation is the Quantity Rate Analysis which was discussed earlier. In addition to traditional cost estimating approaches, alternative cost estimation models have been developed and explored in recent years in an endeavor to advance the reliability of cost estimates in predicting the actual final costs. Many of the developed methods are either statistical or based on artificial intelligence.

Some of the Non-Traditional methods are:

- Regression Analysis
- Reference Class Forecasting
- Case Based Reasoning
- Neural networks
- Support vector machine
- Fuzzy Inference System
- Monte-Carlo Simulation

4.1. Regression Analysis

Regression analysis is a statistical process for estimating the relations among variables. It convey the relationship between a dependent variable and one or more independent variables. More specifically, regression analysis helps to recognize how the characteristic value of the dependent variable vary when any one of the independent variables is varied, while the other independent variables are fixed. Regression analysis is used for prediction of variables. It is also used to understand which among the independent variables are related to the dependent variable, and to investigate the forms of these relationships. By using RA, the model of the cause and consequence relationship between dependent variable and independent variable(s) can be determined and estimation can be
performed. RA is mainly divided into two groups: linear regression analysis (LRA) and nonlinear regression analysis (NLRA). The purpose of linear regression is to establish the nearby line to the data, while nonlinear regression is a more general technique to fit a curve through the data. According to the number of independent variables, LRA can be divided into two groups: simple and multiple. Simple linear regression analysis (SLRA) is used to estimate the dependent variable with only one independent variable, while multiple linear regression analysis (MLRA) is used to estimate the dependent variable by using two or more independent variables. Regression Analysis could be done using many software. The commonly used among them are SPSS, Minitab, Python, R, Excel etc...[10]

4.2. Reference Class Forecasting

Reference class forecasting is a method of calculating the future by means of similar past situations and their results. Reference class forecasting predicts the outcome of an intended action based on actual results in a reference class of similar actions to that being forecast. Reference class forecasting was actually developed by Daniel Kahneman and Amos Tversky, but it was first applied by Flyvbjerg on large British transportation projects.

In addition to possible technical or political explanations for cost overruns, RCF interprets inaccuracies in estimates as being psychological because of optimism bias. To deal with the problem, RCF suggests that when making the first decision about a project, whether to go ahead or not, it is recommended to take the outside view rather than the inside view as the former is much more likely to produce an accurate forecast. The outside view means that instead of concentrating on the present project, the planner should focus on past experiences on similar projects, referred to as the reference class.

Reference Class Forecasting is a method for systematically taking an outside view when planning projects, by basing forecasts on actual performance of comparable projects rather than focusing only on the project in progress. The American Planning Association has recommended that planners shall not only rely on the inside view forecast technique, but instead use Reference Class Forecasting in addition to traditional methods as a way to improve accuracy.

To perform RCF, four steps are required:

- Gathering the project’s planned and actual data: The collected data including the planned and actual data that are gathered for grouping into different classes.
- Identifying the reference class (es) from past projects: It is important that the class should be statistically meaningful and should be truly comparable with the specific project.
- Establishing a probability distribution for each reference class based on the gathered data: This requires entry to possible data on cost overrun for a number of projects within the reference class to make a statistically momentous conclusion.
- Compare the specific project with the reference class distribution: After establishing probability distribution, the most likely outcome for the particular project is established.

[8]

4.3. Case- Based Reasoning

Case-based reasoning (CBR) could be defined as the process of solving new problems based on the solutions of similar past problems. For example, An auto mechanic who fixes an engine by recalling another car that exhibited similar symptoms is using case-based reasoning. CBR is not only a dominant method for computer reasoning, but also a method for computer reasoning. CBR solves the new problem by adapting previously determined solutions of the similar previous cases and storing the new successful solution for future use.

The basic idea behind CBR is the hypothesis that similar problems have similar solutions. An aim of constructing a case-based system is to use the notion of similarity that best fits with this hypothesis. Generally, the CBR problem-solving process has four steps and is known as the four Rs of CBR. They are:

- Retrieve: Given a problem, retrieve from memory cases relevant to solving it. A case consists of a problem, solution, and annotations about how the solution is derived.
- Reuse: The solution is mapped from the previous case to the target problem. This may involve adjusting the solution as needed to fit the new situation.
- Revise: After mapping the previous solution to the target situation, test the new solution in the real world and, if necessary, revise.
- Retain: After the solution has been successfully adjusted to the target problem, store the resulting experience as a new case in memory.

The four R’s of CBR could be represented pictorially in the form of a cycle as:
It has also been observed that CBR can enhance the accuracy of construction cost estimates. However, there are challenges related to the process of retrieving knowledge and information that still need to be addressed.

4.4. Artificial Neural Networks

Artificial Neural Networks (ANN) are also referred to as connectionist systems. NN is a computer system that simulates the learning process of the human brain based on a simplified model of the biological neurons in the human brain and the relations between them. Neural Networks are based on a large collection of neural units loosely modeling like brain solves problems with huge clusters of biological neurons connected by axons. Each neural unit is connected with many others, and links can be enforcing or inhibitory in their effect on the activation state of connected neural units. Each individual neural unit may have a summation function which combines the values of all its inputs together. An artificial neural network is modeled in a mathematical manner to implement an intelligent form as shown in the human brain, for utilization in engineering or in other field. Basically, the network consists of several layers, including an input layer, a hidden layer, and an output layer, and each layer contains neurons. Neurons determine the optimum value through a summation and transfer function. The set of inputs, which is the outputs from another neuron in input layers, are delivered by neurons. Each input data is multiplied by the connection weight, and then the weighted inputs provide output value, which is modified by the transfer function.[10]

4.5. Support Vector Machine

Support Vector Machines (SVMs, also called support vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Support vector machine (SVM) is a learning theory developed by Vapnik. SVM that has two main categories, support vector classification and support vector regression. In particular, in the model constructed using SVR, the goal is to find a function $f(x)$ that has at most $\varepsilon$ deviation from the actually obtained target value $y_i$ for all the training data, and is simultaneously as flat as possible. SVM is one of the most popular supervised artificial intelligence methods and widely used for classification and clustering purposes. There are several SVM available in webpage for academic and commercial purposes, such as, LibSVM, LS-SVMlab, SVMlight and so forth. However, LibSVM is used in this work as it has library which could be easily integrated to Matlab, R and other programming languages interfaces with consideration for real implementation. In general, SVM is also extended to solve regression problems, and thus support vector regression is applied to solve non-linear regression problems by mapping non-linear regression problems to linear regression. [10]
4.6. Fuzzy Inference System

Construction applications are increasingly using computer modeling techniques to help make decisions and estimate costs, performance, quality or time. Fuzzy set theory and fuzzy expert systems are used increasingly in situations where little deterministic data are available. Fuzzy logic was invented by Zadeh in 1965, and has got many applications in many sectors of industry. "Fuzzy set theory was originally devised to model uncertainty associated with human perception or subjective probability judgments". Fuzzy set theory has been used for construction management applications such as risk assessment and pricing construction risks. It can also be used for project control issues such as scheduling estimating project network analysis cash flow analysis evaluating alternative construction technology, crane selection and assisting in selecting corrective actions when problems arise on the construction site etc...

Fuzzy control is the most important application in fuzzy theory. Using a procedure originated by Ebrahim Mamdani in the late 70s, three steps are taken to create a fuzzy controlled machine:

- Fuzzification (Using membership functions to graphically describe a situation)
- Rule evaluation (Application of fuzzy rules)
- Defuzzification (Obtaining the crisp or actual results)

[18]...

Fig 4: Steps in fuzzy system

4.7. Monte-Carlo Simulation

It is the representation of real world by numbers and other symbols. Monte Carlo Simulation is defined as the numerical Technique for conducting experiments on a digital computer which involves certain types of mathematical and logical relationship necessary to describe the behavior and structure of complex real world system over extended periods of time.

The Monte Carlo simulation is widely used for prediction of response or output of complex systems in almost all areas of research interest, including construction management. The Monte Carlo simulation can generate dependent random variables with designated correlation structures and, thus, fit a probability distribution (either parametric or empirical) and/or examine confidence levels of simulated results or outputs (e.g., total project cost). For an effective Monte Carlo simulation, the marginal distribution function of every random variable and dependence structure between them is needed. There are only a very limited number of parametric multivariate pdfs, e.g., normal, lognormal, beta, and t-student. When historical data is available, a parametric marginal distribution can be fitted to these data, using maximum likelihood or method of moments, and subsequently goodness-of-fit of the fitted distribution can be verified statistically. In the absence of reliable historical data, appropriate marginal distributions can be established subjectively based on experts’ experience. However, in any case, there is no guarantee that all marginal follow the same parametric distribution family. In the Monte Carlo simulation, copulas are a very useful tool for modeling dependence of random variables. While the Gaussian copula is the traditional choice for modeling dependence, it is most sensitive to the center of the distribution, which implies tail independence. In other words, an important reason to consider other copulas than Gaussian copula is its failure to capture dependence between extreme events/values. These extreme values are much more important from risk analysis point of view (i.e., cost and time overruns). Clearly selection of other types of copula should be based on their robustness criteria and upon testing their adequacy in modeling specific data sets.[2]

3. CONCLUSIONS

Realistic estimation of construction cost is vital for both successful planning and completion of every construction project. Also in the face of uncertain conditions, reliable cost forecasts become an important source of information for decision making by all construction parties. In order to estimate the construction cost accurately several Non-traditional methods are employed. In this study some of the non-traditional methods along with their application are reviewed based on various literature study. This study may help to encourage local contracting companies to do estimation for project cost using Non-Traditional methods and software packages.

ACKNOWLEDGEMENT

Authors praise the Almighty for showering his blessings without which we could do nothing and granting the capability to proceed successfully. Authors like to express sincere gratitude to the principal Dr. V H Abdul Salam for providing such quality facilities in our esteemed institution for our paper. Authors extend sincere gratitude to all teaching and non-teaching staffs for their valuable advice and help to move forward with the paper.
REFERENCES


Cooperative Organization Sponsored Jointly by the Virginia Department of Transportation and the University of Virginia).


