# Analysis of Cost Overrun in Highway Construction Projects Using Multiple Regression and Artificial Neural Networks

**Thaseena. T[1], Vishnu .K[2],**

[1]Student, Dept. of Civil Engineering, Sri Vellappally Natesan College of Engineering, Kerala, India
[2] Assistant Professor, Dept. of Civil Engineering, Sri Vellappally Natesan College of Engineering, Kerala, India

---***---

**Abstract** - *Cost overruns are more common in infrastructure projects especially, more common in road construction activities. There existed a need to develop a probabilistic cost overrun analysis model in construction projects as a decision support tool for contractors before the bidding stage. The objective of this study is to identify the critical factors affecting cost overrun and obtain statistical models using multiple regression and artificial neural networks. Regression models are obtained using SPSS software. Using MatlabR2011b a neural network models are obtained in which training of the neural network model is administrated via back propagation algorithm. The model is then validated and results show the better analysis of cost overrun in highway projects.*

***Key Words*: Cost overrun, regression analysis, neural networks, training, testing.**

## 1. INTRODUCTION

The construction industry of India is an important indicator of the development as it creates investment opportunities across various related sectors. The construction of large projects such as bridge, road, etc. comes under heavy construction category. The construction works that involve buildings of real estate ones such as residential or commercial real estate assets, etc. include in general category.

Several reasons cause cost overrun which includes changes in estimate design scope etc. A common reason for cost overruns is the inaccuracy of cost estimates. When the bids for subcontracts or the actual costs come in, they are often higher than anticipated. Such cost overruns are due either to incorrect estimates or to changed conditions in the marketplace. Review can do for cost estimates before placing orders to identify mistakes or changed conditions. An overall review may find that increases in some areas are compensated by decreases in others. Clients are able to adjust requirements to reduce costs or seek out lower-cost suppliers. Advising the business owners or managers of possible higher costs at this stage gives them the option of making changes and maintaining their budget.

Sometimes, the designs or drawings that form the basis of the project are not realistic. There may be a combination of specified features is difficult to achieve or that drawings show an incorrect arrangement. Executing the project as specified will either cost extra or cause problems that must be resolved later at additional cost. A project manager should be continuously comparing plans with executed work to find such discrepancies early and correct them.

Kerala has narrower highways compared to other parts of India. National highway authority of India upgrades highway to a minimum four lanes, 60 meters wide, grade separated highway etc. But in Kerala national highways are being upgraded as only 45 meters wide highways. The reasons for this are difficulty and disinterest in land acquisition. Many highway projects widening and maintenance had stopped working due to the protests from local political parties. This leads to the poor conditions of roads in Kerala. As a result the project costs increases day by day.

### 1.1 Need for the Study

Cost variations occur in construction industry due to change of material prices, economy, labour productivity, and delay in repair and maintenance. In Kerala the conditions of the roads are very poor. The road repair works are kept pending which results in the formation of potholes leads to troublesome journey. The day by day increase in material prices and labour prices leads to increase in construction cost.

### 1.2 Scope

This study aims a quantitative approach in identifying and assessing the various factors affecting the cost overrun in road projects in Kerala, by conducting a questionnaire survey among various engineers, clients and contractors. Further to develop multiple regression and neural networks models for cost overrun estimation in construction project.

### 1.3 Objectives

The various objectives include:

- To identify the common and relevant factors causing cost overrun in road projects.
- To conduct a questionnaire survey among consultants, contractors and engineers.
- To develop statistical models to assess the significant factors affecting cost overrun using multiple regression and neural network method.

- To validate the results obtained from the statistical models.

## 2. LITERATURE REVIEW

Al-Momani [1] developed a quantitative regression model for estimating the actual time using the data of 130 public building projects constructed during 1990-1997 in Jordan. It was concluded that the main causes of delay in construction projects are caused by designers, owner changes, weather, differing site conditions, delays in material deliveries, economic conditions and increase in quantities.

Ayman *et.al* [2] proposed a probabilistic to predict the risk effects on time and cost of construction projects. Project managers and consultants can use the model in estimating project cost and duration based on historic data. Statistical regression models and sample tests are developed using real data of 140 projects. The predicted proposed models are accurate with a probability of 95%. The goodness of the fit is represented by the calculated R-squared value of 99.92%. This means that 99.92% of the variance in the variable is explained by the model.

Elbeltagi *et.al* [3] investigated factors that significantly influence highway construction and an artificial neural network model is developed for predicting the cost. The network is trained and tested with a total of 67 projects historical data. Training of the model is administered via back-propagation algorithm. The results obtained from the ANN model with two layers containing 5 neurons, and 30 neurons on the first and second layers, respectively, were consistent and gave values of predicted unit cost very close to the actual unit cost. The percentage error for the training phase is 0.034%, for the testing phase is 3.058% with total average error is 1.55%.

## 3. METHODOLOGY

The detailed methodology adopted in this study is explained below. The methodology includes the identification of factors affecting the cost overrun in road projects and to conduct a questionnaire survey. Further this data is used to develop regression models and neural network models to predict the cost overrun.

### 3.1 Identification of Various Factors Affecting Cost Overrun

As the initial step previous research papers were reviewed to investigate various factors causing cost overrun. 45 major factors were identified which include terrain conditions, design problems, problems among construction contractors, road conditions etc.

### 3.2 Preparation of Questionnaire Survey

Various literatures were studied and based on the preliminary investigation conducted at the outset of this study. 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 overruns. The respondents were asked to rank the individual effect of cost overruns 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 no effect-0, low effect-1, medium effect-2 and high effect-3. The questionnaire survey was carried out among three groups of construction practitioners which are architects, engineers and contractors related to road construction. Questionnaires were distributed in different firms and 35 were returned as response.

### 3.3 Identification of Critical Factors

After conducting the questionnaire survey the data was analyzed and the critical factors affecting cost overrun were identified using Relative Important Index (RII) method. A four point Likert scale of 0-3 was adopted to assess the degree of agreement of each cause or scale of impact. Each factor was given a scale of 0 to 3, so that person could easily express the severity range or impact. i.e. 0 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 | |
|---|---|
| 0 | No effect |
| 1 | Low effect |
| 2 | Medium effect |
| 3 | High effect |

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

$$RII = \frac{\sum W}{(H \times N)}$$

Where **∑W** is the total weight given to each factor by the respondents, which ranges from 0 to 3 was calculated by an addition of the various weightings given to a factor by the entire respondent , H is the highest ranking available (i.e, 3 in this case )and N is the total number of respondents that answered the question. Based on the RII value the factors were ranked and identify the most critical factors. The six critical factors were selected as input for both multiple regression and ANN models for prediction of cost overrun.

### 3.4 Cost Estimation Using Regression Technique

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 or predictors. By using the critical factors and cost data a regression model was modeled using SPSS software. The model developed is used to predict the cost overrun in road projects.

### 3.5 Neural Network Method

Artificial Neural Networks (ANN), as one of the artificial intelligence techniques, has been extensively used for cost estimate. ANN presents itself as an approach of computation and decision making that may potentially resolve some of the major drawbacks of traditional estimating technique. The major strength of ANNs is their ability to learn from examples and to generalize that knowledge to novel cases. Using the critical factors and cost data a neural network model was developed in MATLAB. Based on an appropriately configured ANN model and a sufficient set of past completed highway projects, an ANN model was able to arrive at accurate forecasts of the cost of a new construction highway projects.

Data are generally normalized for effective training of the model being developed. The normalization of the data is the scaling of the input and output pairs within the range (-1, 1) or the range (0, 1) depending on the processing function. It is used to allow the squashing of the values to improve the network performance [4]. Furthermore, the neural networks usually provide improved performance when the data lie within the range (0, 1) because a sigmoid function is used, a slow rate of learning occurs near the end points of the sigmoid function.

Hidden layer is a layer of neurons in an artificial neural network that does not connect to the outside world but connects to other layers of neurons. Number of hidden layers is determined by finding the network structure that generates the minimum RMS value for the given problem output parameters. The network architecture consists of ten inputs i.e. critical factors obtained from questionnaire survey and one output i.e. cost overrun.

The training function used in this model is Levenberg – Marquardt (TRAINLM). 75% of the data was used for training the network. Performance algorithm used here was mean squared error. Feed forward back propagation was used as the training algorithm.

## 4. RESULTS AND DISCUSSION

About 45.7 % were responded as the poor communication between construction parties had high effect in cost overrun in road projects whereas 54.28% people were responded as medium effect. Mistakes in design had high effect declared by 57.14% people whereas only 31.4 % declared as have only medium effect. For the case of late submission of nominated materials 51.42% were responded as medium effect and 45.7% were responded as high effect. About 60% were responded have medium effect for poor workmanship where as 40% responded as high effect.

### 4.1 Critical Factors

Based on the ranking ten major critical factors were obtained are late issuing of approval documents, poor workmanship, poor communication between construction parties late submission of nominated materials, mistake in design, delays in decision making, payments delay, availability of materials, limited construction area, design changes which had an RII value greater than 76%.

Late issuing of approval documents had an RII of 82.85% which is ranked as 1. For the government projects there will be delays for issuing documents from different authorities there by increasing the material costs and cost escalation will occurs. Poor workmanship is ranked as two since it has got an RII of 81.9% represents the availability of skilled and efficient workmen in road projects are also a major criteria.

Poor communication between construction parties, late submission of nominated materials and mistake in design had got an RII value of 80.95% and was indicated as rank 3. The impact of poor communication between individuals causes various conflicts and which delays the construction work. As the delay in submission of material increases cost also increases. Mistakes in design also play an important role in cost overrun. To overcome this problem proper design according to specifications should adopt. Improper planning, inadequate site investigation, misinterpretation of data, unaware of future needs are some of the causes for mistake in design. So proper planning, adequate investigation of site and accurate design procedure are needed to execute the project with high precession.

Delays in decision making and payment delay are ranked as four had an RII of 79.04% among the various factors. Almost 95% of road projects are funded by Government organizations and due to various Government policies and reasons, funds will not be given in time. This will reflect in the performance and progress of works. Due to the delays in payment, contractor will not able to circulate the money and his cash flow is mainly affected. Changes in objectives or plan results delay in taking decisions.

Availability of material is ranked as five. During the periods when the levels of activity developments are unusually high in particular region, there might be chance shortage of materials used in the construction activities. If this is not anticipated in the planning stage itself, it will lead to increase in cost due to transportation of materials from far away areas. So provisions are given at contract at the stage of original cost estimate for avoiding the delays. Limited construction area and design changes are ranked as six had an RII value of 76.19% since the availability of land condition are very poor in Kerala it will affect construction cost and land acquisition cost.

The least affecting factors were identified are monopoly, inflation, exchange rate of fluctuation etc. Monopoly had an RII value less than 45.71%. Inflation got an RII of 46.67% and exchange rate fluctuation had an RII 42.85%.

### 4.2 Regression Models

The common static associated with regression analysis is $R^2$ which is the deviation of the actual value of the dependent variable from the regression line. It is equal to one minus the ratio of the sum of squared estimated errors to the sum of squared deviations about the mean of the dependent variable. Intuitively, the sum of squared deviations about its mean is a measure of the total variation of the dependent variable. The sum of squared deviations about the regression line is a measure of the extent to which the regression fails to explain the dependent variable. Hence, the $R^2$ statistic is a measure of the extent to which the total variation of the dependent variable is explained by the regression.

**Table -2:** Cost overrun model

| Model | R | $R^2$ | Adjusted $R^2$ | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | 0.989 | 0.978 | 0.870 | 1.41368 |

The table 2 shows the model summary in which correlation coefficient R=0.978 which represents the entire model is highly correlated with cost of the project. The total variation is defined by all the variables are 87% that is the project is feasible from data point of view with only 13% loss of data. 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 = 0.989 which is nearly equals to 1, hence the correlation can be found out with very less amount of loss in data.

**Table-3** ANOVA statistics

| ANOVA | | | |
|---|---|---|---|
| Model | Sum of squares | Mean square | Sig. |
| Regression | 180.115 | 18.012 | 0.014 |
| Residual | 3.997 | 1.998 | |
| Total | 184.112 | | |

The model equation obtained

$$Y = -31.797 + 4.322\ LIAD - 11.459\ PW - 0.387\ PCBP + 3.470\ LSNM + 2.707\ MD + 7.676\ DDM + 1.109\ PD + 1.509\ AM + 4.955\ LCA + 4.946\ DC$$

The predictors of the model are the late issuing of approval documents(LIAD), poor workmanship(PW), poor communication between construction parties (PCBP), late submission of nominated materials (LSNM) , mistake in design(MD), delays in decision making(DDM), payments delay (PD), availability of materials(AM), limited construction area(LCA) and design changes(DC).

The Analysis of variance or ANOVA table shown in table 3 represents the acceptability of the model from statistical perspective. This table indicates that the regression model predicts the dependent variable significantly well. The regression row display information about the variation accounted for by the model. At the "Regression" row the "Sig." column indicates the statistical significance of the regression model.

The regression model was obtained is then validated using 10 projects.

**Percentage cost overrun = ((Actual cost – Estimated cost)/Actual cost) *100**

The actual cost overrun vs predicted cost overrun was shown in chart-1.

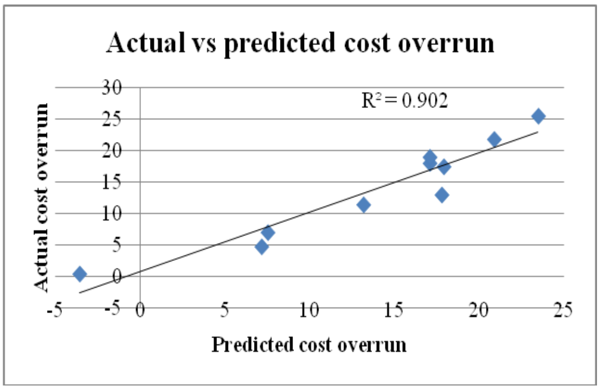


**Chart-1** Actual cost overrun Vs predicted cost overrun

### 4.3 Neural Network Models.

By using the ten critical factors as the input and cost overrun as the output neural network models was developed in MatlabR2011b. The network type used was feed forward back propagation algorithm and the performance function is used as mean squared error. Network 10-13-1 shows the best performance for the correlation value of about 0.9923 which means the network 10-13-1 is the best architecture which was shown in figure 1

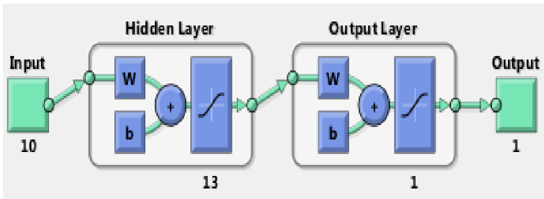


**Fig -1** Best ANN architecture 10-13-1

Here the training data indicates a good fit. The validation and test results also show R values that greater than 0.9. Each pattern presentation represents an iteration and the presentation of entire of the entire training data set is called an epoch. The weight changes should be computed from mean square error entire training set keeping the neural network fixed. The training and testing results shows correlation value of 0.9997 and 0.9967 respectively. The correlation value obtained from training represents high goodness of fit was indicated in Fig-2 .

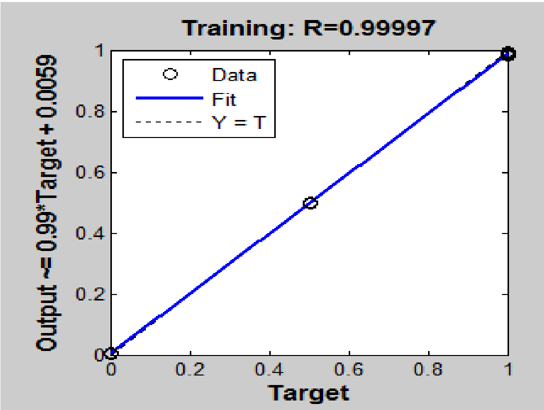


**Fig 2-**Correlation for ANN 10-13-1

The scatter plot is helpful in showing that certain data points have poor fits. If there is a data point in the data set whose network output is close to 0.6, while the corresponding target value is about 0.5.

The overall correlation value obtained for traing testing and validation is 0.99239 which shows the neural network model is more consistent than regression model.

Validation of data was carried out by NNtool itself. 10% of data is used for validation. During iterative training of a neural network, an epoch is a single pass through the entire training set, followed by testing of the verification set. The correlation value for validation was obtained as 0.95318 which is greater than 0.9 which represents the best validation performance for the data obtained.

## 5. CONCLUSIONS

Cost overruns occur in every construction project and the magnitude varies considerably from project to project. Only some projects are being completed within the budget. So it is essential to define the actual causes of cost overrun in order to minimize the impact of the increase in cost in any construction project. The total variation is defined by all the variables are 87% that is the project is feasible from data point of view with only 13% loss of data. In neural network models the best performance of architecture were obtained was 10-13-1 with 10 input layers and 13 hidden layers with one output which is cost overrun which has a correlation value of 0.99239 shows the best statistical model.

Future studies could be performed for different specific types of construction projects, such as railway construction projects, building housing projects, utility projects, viaducts and dam construction projects, etc using neuro fuzzy models and genetic algorithm models. Extension to the feed forward algorithm includes new type of architecture such as recurrent neural networks, radial basis function networks and recursive neural networks which can be used.

## REFERENCES

[1] H.A. Al-Momani, Construction Delay: A Quantitative Analysis. International Journal of Project Management, vol. 18, 2000, pp. 51-59.

[2] 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, vol. 2, 2008, pp. 250-266.

[3] E. Elbeltagi, O. Hosny, R.Abdel-Razek, and A. El-Fitory, "Conceptual Cost Estimate of Libyan Highway Projects Using Artificial Neural Network," International Journal of Engineering Research and Applications, vol.4 , 2004, pp. 56-66.

[4] T.Hegazy, O. Moselhi and P.Fazio, "Developing Practical Neural Network Application Using Backpropagation," Journal of Microcomputers in Civil Engineering, Vol. 9, No. 2, 1994, pp. 145-159.