ISSN 2249-5908
Issue 6, Vol. 5 (September 2016)

Analysis of Construction Cost Overrun in Residential Building using Statistical Fuzzy, Multiple Regression and PLS-SEM Method

Gayathri Thampi[#1], Lekshmi M. G[#2]

#1 Karukayil house, Pallipad P.O, Haripad, Alappuzha, 9497874210,
#2 Madan villa, Cherukunnam, Thekkekkara P. O, Mavelikkara, 9446028530,

ABSTRACT

This paper aims at identifying the factors causing cost overrun in residential building using three methods. A list of 55 factors contributing to cost overrun was prepared based on literature survey and ranked using relative importance index (RII) scale based on the opinion of the engineers by distributing the list in the form of questionnaire to 30 respondents. This paper presents an application of three methods to analyse and develop model for predicting cost overrun in residential building. Statistical fuzzy analysis was performed in fuzzy logic toolbox of MATLAB Programme Software, regression analysis using SPSS and partial least square structural equation modeling (PLS-SEM) using SmartPLS software. Finally, the results has been validated and compared to find out the most reliable and accurate method of analysis.

Key words: cost overrun, relative importance index, fuzzy, PLS-SEM, SmartPLS, SPSS.

Corresponding Author: Gayathri Thampi

INTRODUCTION

Construction industry is a very important industry that plays a vital role in the socioeconomic growth of a country. Economically, it contributes to overall GDP of a country. It improves the quality of life by providing us infrastructures such as roads, hospitals, schools and other basic and enhanced facilities. Hence, it is fundamentally crucial to make the construction projects to complete successfully within time, budget and quality expected. But construction industry faces many problems such as low quality, low productivity, cost overrun, time overrun etc. of these, cost overrun is the major problems as money us always of greater importance.

Cost overrun is a global phenomenon in the construction industry and very rarely projects are finished within the budgeted cost. The issue of cost overrun in construction industry us very dominant in both developing and developed countries but this trend us very severe un developing countries where these cost overrun sometime exceeds 100% of the anticipated cost. Also the accuracy of early cost estimation in engineering and construction projects is extremely important to both owners and project teams. Decision making in the early stage of project has a significant impact on the project. This study aims to analyse the cost overrun factors in residential building using three methods. The building factors which affect cost overrun were surveyed. A questionnaire survey and relative importance index ranking technique were used to indicate the most important five factors. Then a partial least square analysis will be conducted to identify the most important factors contributing to cost overrun. And a statistical fuzzy and multiple regression were used to develop models. Finally, a comparative study was conducted to identify the best method among the three which will give an acceptable degree of accuracy.

MATERIALS AND METHODS

This study us conducted through the following steps:

1. A literature review was carried out to cover the previous studies related to cost overrun and thus the factors contributing to cost overrun and thus the factors contributing to cost overrun was identified.
2. Based on the identified factors, a questionnaire survey was conducted.
3. The identified factors of cost overrun were ranked based on their relative importance index (RII).
4. Identified factors are analyzed using statistical fuzzy, multiple regression and partial least square method.
5. The results obtained in these three methods are validated and compared to find out the most accurate method.

DATA COLLECTION

Data collection was carried out by conducting questionnaire for identifying the significant factors affecting cost performance. The questionnaire includes two parts, part A and part B. Part A includes project details and part B includes the severity, frequency and degree of agreement to each factors. A five point likert scale was used to assess the degree of agreement of each factor. This five point likert scale was converted to a relative index value using the formula below:

$$RII = \sum W \div H \times N$$

Where ∑W is the total weight given to each factor by the respondents, which ranges from 1 to 5 and is calculated by an addition of the various weightings given to a factor by the entire respondents, H is the highest ranking available and N is the total number of respondents that have answer the question. In this study 8 groups of 55 factors are identified. Table 1 shows the 8 groups of factors of cost overrun and their RII value and corresponding rank.

Group	Description	RII	Rank
1. Contractor's site management related factors	➢ Traffic control and restrictions at site	0.520	20
	➢ Inaccurate time and cost estimates	0.487	23
	➢ Site accidents	0.180	46
	➢ Differing site conditions	0.167	48
	➢ Mistakes during construction	0.153	49
2.Design and documentation related factors	➢ Changes in drawings	0.540	15
	➢ Design errors	0.527	19
	➢ Ambiguities and inconsistency of specifications and drawings	0.507	21
	➢ Design changes	0.434	27
	➢ Changes in specifications	0.394	30
	➢ Incomplete documents	0.313	37
3.Funancial management	➢ Delay in progress payments by owner	0.727	2
	➢ Inflation	0.654	6

related factors	➢ Financial difficulties of owner ➢ Economic condition ➢ Cash flow problems faced by contractor	0.647 0.614 0.603	7 10 11
4.Owners fault and improper communication related factors	➢ Poor communication ➢ Disputes ➢ Uncooperative owner ➢ Inefficient coordination by owner planning and design changes ➢ Delay in issuance of change order by owner ➢ Delay in settlement of contractors claim by owner ➢ Excessive bureaucracy in owner and administration ➢ Delay in furnishing and delivering the site to contractor ➢ Suspension of work by owner ➢ Political influence ➢ Delay in making decision by owner ➢ Inference by owner in construction operation phase	0.554 0.313 0.494 0.133 0.113 0.220 0.300 0.267 0.387 0.153 0.334 0.667	15 37 22 50 52 43 38 40 31 49 35 3
5.Human resource related factors	➢ Shortage of manpower ➢ Low skill of manpower ➢ High cost of labor	0.327 0.534 0.647	36 18 07
6.Non human resource related factors	➢ Change in material price ➢ Effect of subsurface condition ➢ Delay in material procurement ➢ Imported materials ➢ Shortage of materials ➢ Failure of equipment ➢ Shortage of equipment	0.794 0.120 0.580 0.474 0.667 0.480 0.374	1 51 32 24 5 23 32
7.Project management related factors	➢ Change in scope of project ➢ Poor coordination by contractor with parties in project ➢ Poor coordination by consultant engineer ➢ Delay in mobilization to start the project ➢ Negotiation during construction ➢ Underestimation of productivity ➢ Ineffective control of project process by the contractor ➢ Delay in performing inspection and testing by the engineer ➢ Mistakes during construction ➢ Difficulties in obtaining work permits from the authorities	0.367 0.487 0.620 0.414 0.253 0.634 0.167 0.460 0.580 0.720	33 23 9 28 41 8 48 25 13 3
8.Contract administration related factors	➢ Original contract duration too short ➢ Subcontractor problem with contractor ➢ Low performance of the lowest bidder contractor ➢ Spending sometime in finding a suitable	0.674 0.200 0.547 0.587	4 44 16 12

ISSN 2249-5908
Issue 6, Vol. 5 (September 2016)

	contractor		
	➢ Ineffective planning and scheduling of the project	0.494	22
	➢ Delay in field survey done by the contractor	0.180	46
	➢ Interfere with other contractors work	0.440	26

FUZZY ANALYSIS

To develop the model, following steps are performed on fuzzy logic toolbox of MATLAB.

1. Through the analysis based on the RII value the five top most factors have been selected and used in the fuzzy logic analysis to predict the percentage of cost overrun in the construction industries.
2. Construct a five input and an output system in FIS editor. The identified cost overrun factors and cost overrun are considered as the input and output.
3. Membership functions are defined for each of the input and outputs.
4. Rules which connect input to output variables are then defined here total of 25 rules are constructed.
5. Relative importance indices are given as the weightage to each of the fuzzy rules.

Fig 1 shows the rule viewer of the system.

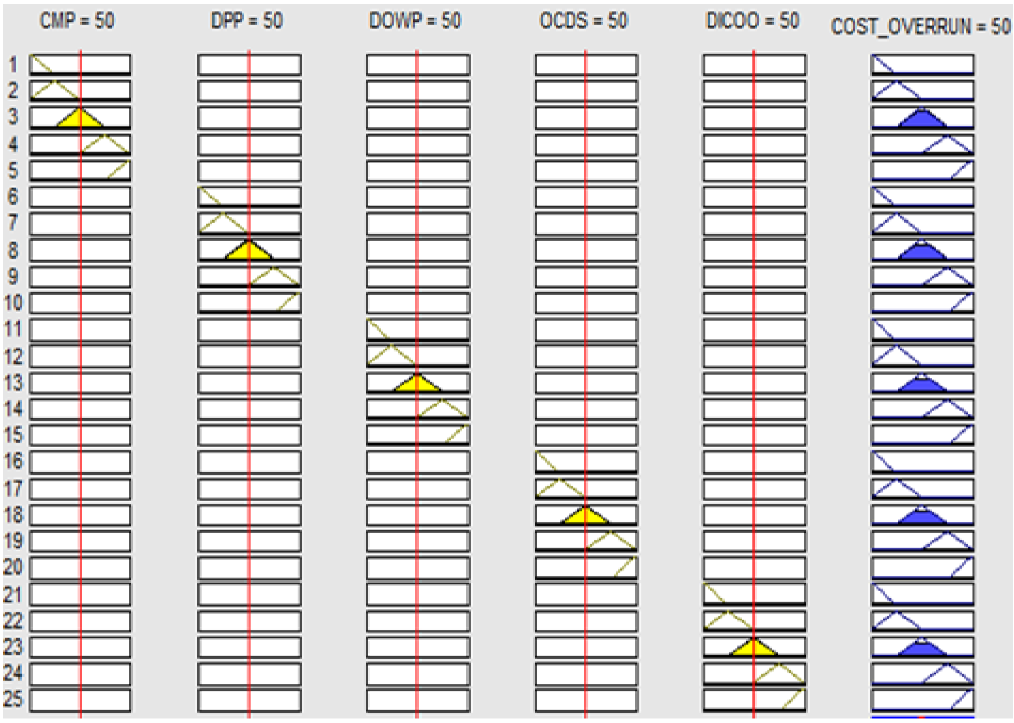

Fig 1: Rule view

To display the dependency of one of the outputs to any one or two of the inputs, it gives surface view of the system. Fig 2 shows the variation of delay in progress payments by the owner (DPP) and change in material prices (CMP) with respect to cost overrun. Similarly variation of cost overrun for different combination of input variables can be obtained.

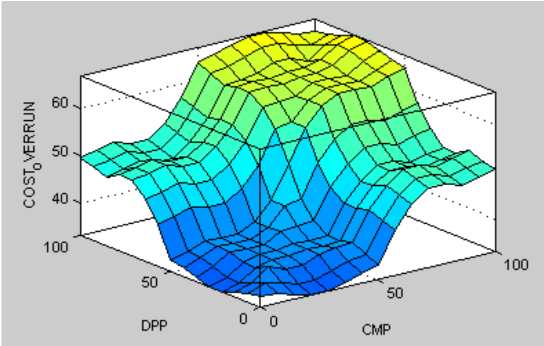

Fig 2: variation of delay un progress payments by the owner and change in material prices with respect to cost overrun

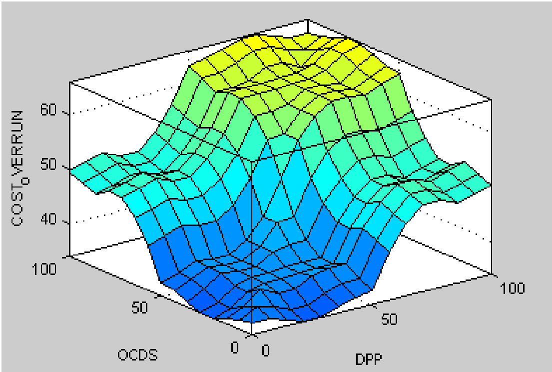

Fig 3: variation of difficulties in obtaining work permits from the authorities concerned and original contract duration too short with respect to cost overrun

MULTIPLE REGRESSION ANALYSIS

In this study multiple regression analysis was carried out in SPSS version 16. Both backward and forward regression analysis can be performed. From literature review, backward regression is more valid and applicable than forward regression analysis. This may be due to the fact that backward regression has the advantage of looking all the variables in the early stage of analysis.

Table 2 and table 3 indicate the final results of backward regression analysis for the project cost overrun and its ANOVA statistics.

Table 2: Model summary

Model	R	R squared value	Adjusted r square	Std. error of estimation
3	0.846	0.715	0.637	0.39473

The R value represents the simple correlation and is obtained as 0.846, which indicates a high degree of correlation. The R^2 value indicates how much of the total variation in the dependent variable, cost overrun, can be explained, which is very large in this case.

ISSN 2249-5908
Issue 6, Vol. 5 (September 2016)

Table 3: ANOVA

Model		Sum of squares	df	Mean square	F	Sig
8	Regression	4.3	3	1.433	9.199	0.002
	Residual	1.714	11	0.156		
	Total	6.014	14			

Table 4 represents the coefficient table which provides the necessary information to predict cost overrun from factors considered. The unstandardized coefficients were used to present the regression equation as:

$$\text{Cost overrun} = 7.487+15.17x_3-11.61x_4+10.332x_5$$

Table 4: coefficients of variables and t test statistics

Model		Unstandardized coefficients		Standardize coefficients	T	Sig.
		B	Std. error	Beta		
3	Constant	7.487	4.399		-1.702	0.117
	DOWP	15.17	6.729	0.493	2.254	0.046
	OCDS	-11.614	5.232	-1.383	-2.220	0.048
	DICOO	10.332	4.191	1.637	-2.465	0.031

PARTIAL LEAST SQUARE ANALYSIS

This method of analysis mainly includes 4 steps:

1. Preparation of theoretical model

   A theoretical model represents the relationship of the identified factors with cost overrun. Fig 4 shows the theoretical model developed.

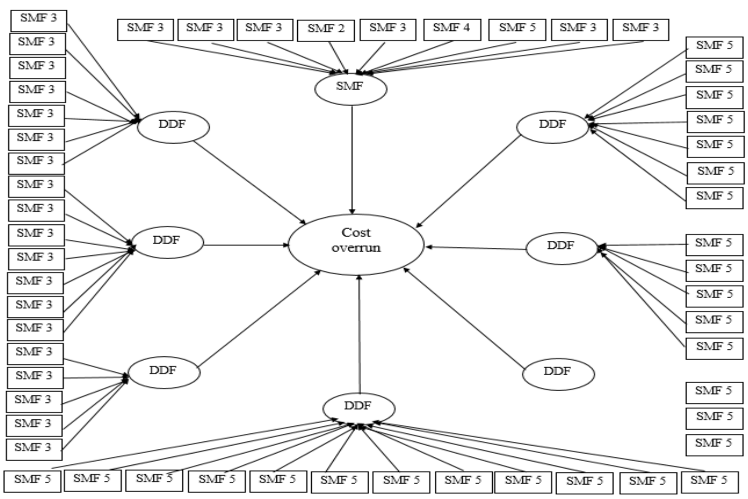

Fig 4: Theoretical model

2. Building the inner model
   Inner model was developed easily in SmartPLS based on theoretical model developed. Fig 5 shows the inner model developed in SmartPLS.

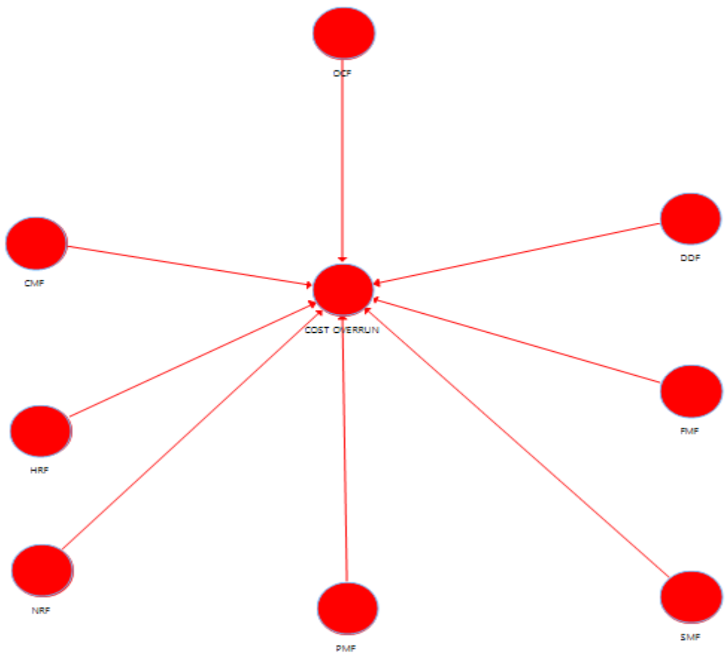

Fig 5: Inner model

3. Building outer model
   Here indicators are linked to latent variables. Fig 6 shows the resulting outer model in smartpls.

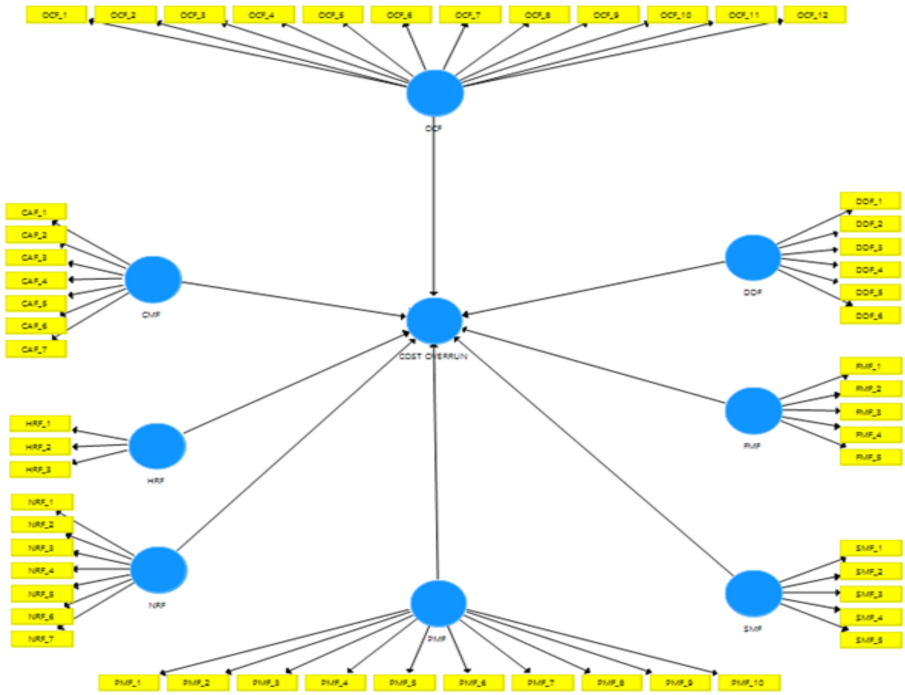

Fig 6: outer model

4. Running the path model estimation

After the indicators and latent variables are linked together successfully in smartpls, the path modeling procedure was carried out.

The evaluation process includes 3 steps:

1. Individual item reliability
2. Structural relationship
3. Overall model fitness
   Overall model fitness for assessing global validity and explaining power of the model. This was done by evaluating goodness of fit (GoF) index. GoF cut-off values for validation of pls models were calculated following the guidelines of Wetzels et al., (2009) suggesting that using 0.50 as the cut-off values for communality and different effect sizes of R^2 as shown in table 5.

Table 5: GoF index and its criteria

GoF	GoF criteria
GoF = $\sqrt{communality} \times R^2$ Range of GoF values: 0<GoF<1	Communality = 0.5 R^2 small= 0.02, medium = 0.13, large = 0.26 Thus, GoF_{small} = 0.10 GoF_{medium} =0.25 GoF_{large} = 0.36

VALIDATION

1. Validation of fuzzy model:
   A detailed case study analysis of 5 building construction projects was carried out to validate the survey findings on most significant factors contributing to cost overrun. These case studies include interviews, discussion with construction managers, detailed study of project documents and contracts. Table 6 presents the details of the projects that have been considered for validation and table 7 shows the impact of five factors in construction projects in terms of loss percentage.

Table 6: Details of project considered for validation

Project details	Project 1	Project 2	Project 3	Project 4	Project 5
Type	Residential	residential	residential	Residential	Residential
Location	Kollam	Kollam	kollam	Kollam	Kollam
Estimated cost	24483000	21886005	21586805	22488000	27352000
Final value	26775358	24884324	24884328	24868800	29834308

Table 7: Validation of fuzzy model

Project details	Project 1	Project 2	Project 3	Project 4	Project 5
CMP	25%	30%	20%	30%	35%
DPP	15%	15%	10%	10%	20%
DOWP	15%	20%	0%	15%	15%
OCDS	0%	0%	0%	0%	0%
DICOO	10%	0%	10%	0%	10%

Cost overrun (estimate)	21.2	26	20.7	26	29.6
Cost overrun (actual)	24.8	29.8	22.9	30.8	32.9
% error = {(actual-estimate) / estimated}*100	16.98%	14.6%	10.63%	18.46%	10.03%

2. Validation of regression model:
Detailed case study analysis of five building construction project, same as considered in case of fuzzy validation, was carried out to validate the survey findings on most significant factors contributing to cost overrun. Table 8 shows the validation of regression analysis.

Table 8: validation of regression model

Project details	Project 1	Project 2	Project 3	Project 4	Project 5
CMP	25%	30%	20%	30%	35%
DPP	15%	15%	10%	10%	20%
DOWP	15%	20%	0%	15%	15%
OCDS	0%	0%	0%	0%	0%
DICOO	0%	0%	10%	0%	0%
Cost overrun estimated	22.657	22.657	17.819	22.657	22.657
Cost overrun actual	24.8	29.8	22.9	30.8	32.9
% error	8.6%	23.67%	22.19%	26.44%	31.31%

INFERENCE

1. PLS SEM method:

The developed structural model of cost overrun factors gives better understanding about the influence of each factor towards the cost overrun generation. The findings of study are summarized as follows:

- Investigated factors have significant effect on residential building cost overrun.
- Developed model is substantial in representing the relationship of the factors on cost overrun with R^2 value of 0.734.
- GoF value of the model was achieved as 0.635 which showed that the model has good explaining power in generalizing the cost overrun problems for residential building.
- Contract administration related factors are major contribution to cost overrun.

2. Regression model:

   The findings are:

   - Among the considered top factors, difficulties in obtaining work permits from the authorities concerned, original contract duration too short and delay in issuance of change order by the owner and changes in material price are insignificant factors.

ISSN 2249-5908
Issue 6, Vol. 5 (September 2016)

- Delay in progress payments by the owner and changes in material price are in significant factors.
- Regression model has a strong correlation coefficient R equal 0.846.
- Coefficient of determination R square equal 0.715 which is a best fit which means that 71.5% of the total variation in cost overrun can be explained by the model.
- Average percent error for each project is not lesser than 20%. Therefore the predicted model in this study will be only an average model to predict the cost overrun that occurs in the construction projects.

3. Fuzzy model:
   The findings are:
   - Different graphs can be plotted to show the variation of different combinations of cost overrun factors with respect to cost overrun
   - Average percent error for each of the project is lesser than 20%. Therefore the predicated model in this study can be considered as the best model to predict the cost overrun that occurs in the construction projects at the planning stage itself and the suitable preventive measures can be adopted to overcome the situations.

CONCLUSION

On comparing the results of analysis using these three methods, fuzzy model was proposed as an effective probability analysis technique in construction projects, since; fuzzy theory is based on uncertainties where there is an inherent impreciseness and it provides mathematical tools to deal with imprecise, uncertain and vague data. Another advantage of this method is the surface viewer. They are types of three dimensional graphical views that can be analyzed by the owner and contractor easily and quickly and can easily understand the variation of cost overrun. And in the validation part, percent error is less than compared to the regression model. On the other hand, PLS-SEM method is an effective in identifying the group of factors contributing to cost overrun.

REFERENCE

[1] S.A.Assaf, B.Al-Khalil, and M.Al-Hazmai, Causes of delay in large building construction projects, J. Manage. Eng., Vol.11, pp. 45-50, 1995.

[2] J.E.Diekmann and M.C.Nelson, Construction claims: frequency and severity, J. Constr, Eng. Manage., Vol.111, pp.74-81, 1985.

[3] T.M.Elhag, and A.H.Boussabaine. Evaluation of construction cost and time attributes, Journal of Management in Engineering, ASCE, Vol.11, pp.55-63, 1999.

[4] B.Flyvbjerg, M.KS.Holm, and S.L.Buhl, How common and how large are cost overruns in transport infrastructure projects? Tranp. Rev., Vol.23, pp.71-88, 2003.

[5] K.Iyer, and K.Jha, Factors affecting cost performance: Evidence from Indian construction projects. International Journal of Project Management, 2005.