

To Study Cost overrun prediction in Construction industry.

A Review

Muhammad Waqar khan¹, Yaseen Mahmood², Muhammad Issa Khan³, Sharifullah Khan⁴

^{1,4}Department of Civil Engineering, Iqra National University Peshawar, Pakistan

²Supervisor, Department of Civil Engineering, Iqra National University Peshawar, Pakistan

³School of Architecture and built environment, Deakin University, Australia

Abstract: In this paper, to predict cost overruns percentage two models were presented in construction industry. In first model 44 factors (based on regression model) were collected from literature which influences construction industry cost performance. 'Contractors' developed a questionnaire survey to know the relative impact of these causes on construction industry projects. Eleven factors were obtained as the most significant causes that lead to cost overrun. For previous fact, occurrence data was gathered for 30 constructions projects which were divided in two sets. 20 projects composed the first set of building model. The results showed strong direct relationship between percentage of cost overrun and the previous 11 causes which highly influenced the cost overrun. These causes are: owner financial condition, contractor cash flow, procurement methods, increase in material cost due to inflation, tender stage competition, currency fluctuations, size of project (small or large), design and approval delays, for quantity variations risk retained by client, detailed drawings, and material estimating inaccuracy. Remaining 10 projects composed as second set, based on case-based reasoning (CBR), and used for validation purposes. CBR is an effective method to solve new problems/issues with experience obtained from past experience of similar project in construction industry. After validation of both models using projects of 2nd set it was concluded that CBR model has less prediction capability as compared to regression model. As attribute weight method gives the highest prediction accuracy of cost overrun percentage after applying absolute value of standardized coefficient (β).

Key Words: Regression Model, Construction industry, Case-Based Reasoning, Cost Overruns, Questionnaire survey

I. INTRODUCTION

In engineering and construction projects accuracy of cost estimates for both project teams and owner is highly essential [1]. In addition, at initial stage of project decision making has played very important role regarding impact on any type of construction project. To evaluate alternatives under a limited definition of scope and constraints, quick and accurate decision making is needed [2]. However, among various factors complex correlation, limited and uncertain information on the project which affect the construction projects cost makes it difficult to predict and manage relevant task [3]. To determine the factors creating construction projects risks several studies have attempted. For large U.S. construction firms, to study the risk attitudes has conducted a survey by [4]. In this survey, among the 23 risk factors composed such as availability of labor, equipment and material, productivity of labor and equipment, design deficiency, changes in scope, various site conditions, safety, contract delayed payment, and work quality were presented as risks with high importance. To allocate risks between contractor and owner [5], the factors related to contract played an essential role. The major factors: influencing contractor contingency decisions i.e. risk rating of country, availability of materials, contract type and payment in advance were explained [6]. For capital projects to predict cost estimate accuracy a multivariate regression model was developed [7]. To solve the issue of predicting construction cost, contingency cost, and cost overrun for construction projects various methodologies was used by previous studies. In the previous studies

some of the methods which were used include: predicting construction cost, statistical methods i.e. multiple regression analysis (MRA) [8-10]. In reconstruction projects for cost overrun predictions regression model was presented [11]. For predicting cost, the contingency models were presented by [7, 12, 13]. • For predicting construction cost repetitive learning methods i.e. artificial neural networks (ANN) represented by [12, 13]. For predicting cost overrun of reconstruction projects presented an ANN model in addition to regression analysis mentioned above was resented by [11].

- Stochastic methods i.e. Monte- Carlo simulation (MCS), [14] conducted a simulation model for predicting the construction cost.

- Analogical methods i.e. Case-based reasoning (CBR) for predicting the construction cost [15-17].

[18] conducted an analysis for a sample of 102 of time and cost overruns educational projects. They concluded that about 32.35% of the selected projects have exposed to cost overrun. For constructing wastewater projects in Egypt to determine the influence ranks of 52 factors causing cost variation based on the quantified relative importance indices a research was conducted by [19]. Four categories were used for factors classifications i.e. Owner originated, Designer originated, Contractor originated and Miscellaneous category. The results were grouped under professional cadre and experience-based group. The study showed that in the project of wastewater construction the owner originated category more affected cost variation caused instead of other three

categories. Lowest bidding related to owner originated category was most significant and predictable factor. Also, he added that because of extra work and bureaucracy in bidding/tendering method, the cost variation can also be made by the owner. Due to foreign firms and aids, the construction industry domination occurred related to "Miscellaneous Category was the less effective factor. On the other hand, CBR having properties that are same to humans' heuristic approach in which experience criteria was used for decisions making.

II. OBJECTIVES

The main objectives of this paper are as following.

- To study/investigate cost overrun of construction industry affected by various factors/causes.
- Two models for prediction of cost overrun percentages in construction projects proposed i.e. regression analysis and case-based reasoning method.

III. SCOPE AND METHODOLOGY

To predict percentage of cost overrun in projects of construction two numbers of models were proposed in this paper. The predictive models are based on regression analysis and case-based reasoning. A methodology of standard nature is adopted. During first step, all past papers were reviews related to the investigation of cost overrun of construction projects causes in order to achieve the objectives of this research paper. The technique of Case-Based Reasoning method was described. During this paper, questionnaire survey was developed including construction projects cost variations/causes to gathered data through developed questionnaire about importance of these causes. In second step, the collected data analyzed in order to achieve most important cost overrun causes. These causes incorporated into predictive models. Demonstration was conducted of Building regression-based model to know how the model predicts project cost overrun percentage a numerical example was prepared. Additionally, in next step CBR model applied to find model performance step by step. After this step, proposed models were validated in this research paper. Finally, the conclusions derived on the bases of results obtained after validation process and comparison of both model's prediction accuracy.

Case Based Reasoning (CBR):

It is the process of studying same nature past cases to solve the new issues by adopting past cases solution and implemented it for new cases issues to strong it for future use successful solution [20]. Four numbers of steps were required by CBR method [2] i.e. representation of case, retrieval of case, adaptation of case, and retaining of case. Cases are represented by attributes describing the situation of issues and its solution. New problems retrieved when past cases matched. To solve the new issue retrieved cases solutions are adopted. Similarly, after approval of new solution(s) of problem retained for use in future.

For calculating attribute similarity: if an attribute is of nominal scale, and previous case value remain same with new case, attribute rated as 01, otherwise rated

zero [3]. On the other hand, if an attribute is either of interval scale or ratio scale, it is scored by Eq. 1.

$$F_{AS} = \begin{cases} \frac{\text{Min}(A_{\text{testcase}}, A_{\text{retricsae}})}{\text{Max}(A_{\text{testcase}}, A_{\text{retricsae}})} & \text{If } F_{AS} \geq MCAS \\ 0 & \text{If } F_{AS} < MCAS \end{cases} \quad (1)$$

Secondly, it is also difficult to assign the attribute weight values that enable to identify the most similar case by an index of corresponding features.

Factors Affecting cost performance in Construction project:

In this paper, 44 factors which cause impact on construction projects cost overrun were identified and gathered form literature review. The data were shown in table 1 (attached at end). For construction projects, in predictive models of cost overrun project these factors act as independent variables.

Questionnaire Survey:

For this study, a questionnaire survey which were divided in two parts was developed to know the importance of construction projects cost overruns causes. The data was collected using these two parts questionnaires. The questionnaire was divided into two main parts. Part first composed of basic information related to respondent's experience, company experience, and company work volume. While second part of consisted, the factors compiled in Table 1 was organized in the form of two priority scaling, one for occurrence frequency, while the other for severity scaling. The occurrence frequency priority scaling was as follows: 5=Always, 4=often, 3=usually, 2=sometimes, and 1=scarcely, while the severity scaling was: 5=very severe, 4=severe, 3=somewhat severe, 2=little effect, 1=very little effect. Based on importance, for each cause the respondent assign scaling for 1 to 5 for both types of priority scaling's. Moreover, the questionnaire also included collection of data for actual past construction projects. The data included occurrence of previous factors impact cost performance of construction projects presented in Table 1 on a yes/no basis. Three conditions to find questionnaire sample size was specified: the precision level, confidence/risk level, and the degree of variability in the attributes being measured [22]. The precision level is the range in which the true value of the population is estimated to be [23], and expressed in percentage points, (e.g., ± 10 percent). For the confidence or risk level, if confidence level of 95% selected, then 95/100 samples have the true population value. The degree of variability refers to the distribution of attributes in the population. The larger size population required to get given precision level if the population is more heterogeneous. To know large and small populations representative sample size Eq.2 and 3 was developed [24].

$$n_0 = \frac{Z^2 pq}{e^2} \quad) \quad n = \frac{n_0}{1 + \frac{(n_0 - 1)}{N}} \quad (3)$$

curve abscissa of that cuts off an area at the tails (1 – equals the desired confidence level,

e.g., 90%), e is the desired precision level, p is the estimated proportion of an attribute that is described in the population q , and is $1 - p$.

The value for Z is found in statistical tables which contain the area under the normal curve. N is the population.

Table. 1 Cost impact factors in Construction Projects

No	Factor Identification	RIW	Rank
F1	Financial condition of the owner	17.4	1
F2	Cash flow of contractor	14.3	2
F3	Method of procurement (open tender or selective tender)	14.2	3
F4	Material cost increase due to inflation	13.9	4
F5	Competition at tender stage(aggressive or not)	12.6	5
F6	Fluctuations in the currency that the payment will be made	11.8	6
F7	Project size (small or large)	11.6	7
F8	Delay in design and approval	11.4	8
F9	Risk retained by client for quantity variations	11.3	9
F10	Drawings (detailed or not)	10.3	10
F11	Inaccurate material estimating	10.3	11
F12	Estimated cost	10	12
F13	Adequacy of quality requirements	9.8	13
F14	Design change	9.5	14
F15	Location of project	9.1	15
F16	How the estimate is prepared? (detailed or not)	9.0	16
F17	Reluctance in timely decision	9.0	17
F18	Difference between low bid and owner's estimate	9.0	18
F19	What is known about the project at the tender stage?	8.7	19
F20	Client characteristics	8.6	20
F21	Unknown geological conditions	8.5	21
F22	Ignorance and lack of knowledge	8.5	22
F23	Liquidated damages	8.4	23
F24	Adequacy of schedule requirements	8.2	24
F25	Conflict among project participants	8.1	25
F26	Quality standards and specifications	7.9	26
F27	Design complexity	7.8	27
F28	Scope change by owner	7.8	28
F29	Time variance	7.8	29
F30	Advanced payment amount	7.5	30
F31	Prequalification of contractors	7.4	31
F32	Level of construction complexity related to new technology	7.4	32
F33	Equipment percentage	7.4	33
F34	Site layout	7.0	34
F35	Time allowed for preparation of estimate	6.9	35
F36	Workload	6.6	36
F37	Contract Type (unit price or lump sum)	6.4	37
F38	Adequacy of dispute settlement procedure	6.4	38
F39	Inspection and testing	6.4	39
F40	Adequacy of safety and environmental requirements	5.8	40
F41	Similar project experience	5.5	41
F42	Weather conditions	5.4	42
F43	Site access	5.4	43
F44	Site congestion	4.4	44

For population of 465 the survey through questionnaire was conducted which represents construction projects contractors work. Due to large population equation 2 was utilized. assumed confidence level= 90% thus $Z=1.65$ from normality tables, assumed $p= 0.5$, e is assumed ($\pm 15\%$). Put values in Eq. 2, initial sample size results $n_0=30.25$. put value of n_0 Eq. 4, $n=28.5$.

IV. RESULTS AND DISCUSSION

Total of 43 numbers of questionnaire were sent to contractors through emails after contacting through telephones and some were given through individual meetings. After this total of 30/43 questionnaire were collected back which showed response rate as 69.76%. For a survey focusing on gaining responses from industry practitioners above rate is considered acceptable [25]. General managers, technical office managers, and construction managers included in this survey. All participants involved in building projects plus additional specializations. While 82% and 42% involved in public water and sewage projects and civil works (bridges, roads, and airports) respectively. The author believes that the variations in positions besides the variations in the specialization for the participants enrich this study to a great extent. This is because data reliability is related to data source and the identification of the position held by the person who completed the questionnaire [26].

Importance Factor (II) was calculated for each cause to identify the causes importance and given by formula as below

$(II) = \text{Occurrence frequency} \times \text{degree of severity}$

Frequency occurrence refers to the probability that any cause (table 1) occurs in a project and contributes to its cost overrun. Above table 1 showed various factors influencing the cost overrun in construction projects in descending order as per their corresponding relative importance weight (RIW) i.e. Highest RIW factor = rank one.

$$RIW = \frac{\sum II * \text{Corresponding Numbers of respondent's}}{\text{Total Numbers of Respondents}}$$

In construction projects, most important factor influencing cost overrun was owner financial condition followed by contractor Cash flow. This clearly showed with negative cash flow for more projects the contractor failed to finance any other project and the project lead to cost overruns due to extension, which leads to cost. Procurement method was ranked as number three which may be open or selective tender. Increase in Material cost due to inflation was ranked 4, since the trend of inflation is probably due to demand exceeding supply, this creates scarcity of goods and hence the prices of materials increase, which result in for construction project cost overrun. On the other hand, [21] found that this factor is among three main causes of cost overrun. Competition at tender stage (aggressive or not) obtained the 05 rank (see Table 1), showed reduction in contingency cost or completely neglected when the

competition is aggressive. Currency Fluctuations ranked 6. Large/ small Project size ranked 7. It seems that cost overrun appears to be less predominant among larger projects instead of smaller projects. Design delay and approval, for quantity variations risk retained by client, drawings (detailed or not), and inaccurate material estimating were ranked: 8, 9, 10, and 11, respectively. In the predictive model, the causes having RIW values less than 10 was not taken in account in order to reduce the number of variables to a manageable number. Table 2 lists the final 11 factors (independent variables) used to develop the regression model.

Regression Based Model:

30 construction projects data were collected which included occurrence of factors (Table 1) on a yes/ no basis, with corresponding actual cost overrun percentage. Two sets of data were formed. The first set contained 20 projects while second set contained 10 projects for the purpose of model building and validation purposes, respectively. An initial experimentation with a regression model that includes all 11 variables using SPSS 13 software was performed by using Forward- stepping and backward-stepping methods. For predicting the percentage of cost overrun this technique gave the same model for construction projects depending on 11 variables (see Table 3) having squared multiple $R= 0.83$. which showed model ability of 83% of data variability and was an excellent indicator of the model's expected performance. The underlying formula of the model is as follows: Percentage cost overrun = $0.214 + 0.046$ (owner Financial condition) + 0.201 (contractor cash flow) + 0.345 (Procurement Method (Open tender or Selective tender)) - 0.177 (increase in material cost due to inflation) - 0.197 (Competition at tender stage (aggressive or not)) - 0.108 (currency Fluctuations) - 0.078 (Project size (small or large)) - 0.284 (design and approval delay) + 0.08 (for quantity variations Risk retained by client) + 0.184 (Drawings (detailed or not)) + 0.08 (estimating inaccurate material). Each of the 11 variables having a 0 (unused), or 1 (used) value.

Table.2 Candidate independent variable Final List

No.	Variable	(RIW)
1	Financial condition of the owner	17.4
2	Cash flow of contractor	14.3
3	Method of procurement (open tender or selective tender)	14.2
4	Material cost increase due to inflation	13.9
5	Competition at tender stage (aggressive or not)	12.6
6	Fluctuations in the currency that the payment will be made	11.8
7	Project size (small or large)	11.6
8	Delay in design and approval	11.4
9	Risk retained by client for quantity variations	11.3
10	Drawings (detailed or not)	10.3
11	Inaccurate material estimating	10.3

Table.3 Regression Model

Constant and Variables	Coefficient
Constant	0.214
Financial condition of the owner	0.046
Cash flow of contractor	0.201
Method of procurement (open tender or selective tender)	0.345
Material cost increase due to inflation	-0.177
Competition at tender stage (aggressive or not)	-0.197
Fluctuations in the currency that the payment will be made	-0.108
Project size (small or large)	-0.078
Delay in design and approval	-0.284
Risk retained by client for quantity variations	0.080
Drawings (detailed or not)	0.184
Inaccurate material estimating	0.080
Squared Multiple R=0.83	

Case Based Reasoning Model:

A case base is developed based on previous cases (first set of projects). Then, those cases which are same to the new cases are retrieved to know percentage cost overrun of the new cases. To retrieve same cases, the similarity values are calculated by multiplying each similarity value ($I_{11}, I_{12}, I_{13}, \dots, I_{1j}, \dots, I_{m1}, I_{m2}, I_{m3}, \dots, I_{mj}$) of each attribute (factor) for a case in the case base and new case by corresponding attribute weight ($W_1, W_2, W_3, \dots, W_j$) and then summing all of them. The weights of attributes are variables. To estimate cost overrun percentage of the new case the case with the highest similarity value is used.

Case Representation and Attributes:

In case of identification of attributes, all attributes (previous 11 factors) were presented which impact construction projects cost included in the regression model. To calculate similarity degree between a new case (test case) and each case in the case base attributes were used. All attributes in this research paper was of nominal scale, so assigned value of one other wise zero. So, If attributes value in each case of case base is the similar as its value in test case assigned one, otherwise, 0. To calculate attributes weight three methods were used : feature counting, MRA standardized coefficient (β), and the standardized coefficient (β) absolute value to enhance the prediction capacity by comparison. To calculate attributes weights by using coefficient (β), MRA run first, in case of BRM. While in method of feature counting with out running MRA attributes weight are determined.

Matching and Retrieval:

The most similar cases will be compared with the new cases in which percentage cost overrun estimated and retrieved the most similar case to the new. Using the mean values if in case base more than one case value similarity value suggested by author.

Adaptation:

For new cases percentage of cost overturn, adaptation was not used in this study because all attributes were of nominal scale.

Case Retaining:

Cases are retained for the purpose of future used in this research study i.e. for knowing solution of the 2nd new case, the 1st new case used among the cases of the case

base. Similarly, for third new case solution, the 1st and 2nd cases were used and so on.

V. CONCLUSION AND RECOMMENDATIONS

In this paper questionnaire survey was utilized for investigating cost overturn causes and their prediction in construction projects. Literature study was done for these causes. These causes were essentially derived from literature. To obtain information related to the occurrence of the previous causes in actual projects on a yes/no basis the questionnaire survey used a structured format. After obtaining results of questionnaires RIW for each cause was calculated in order to calculate its effect on cost performance of project. More ever it was also concluded that all causes having value of RIW greater than 10 are important and incorporated into the model as independent variables. Thus, 11 important causes were obtained. The dependent variable was the cost overrun percentage. To predict cost overrun percentage in construction projects 02 models were developed. The first model based on regression analysis. In Total 30 project, for model building and for validation purpose data 20 and 10 projects were used, respectively. In predicting cost overrun percentage the best model found accurate which contained the previous 11 causes. These are: owner financial condition , contractor cash flow, procurement methods, increase in material cost due to inflation, tender stage competition , currency fluctuations , size of project (small or large), design and approval delays , for quantity variations risk retained by client, detailed drawings, and material estimating inaccuracy.

CBR is the second model used in this paper. Comparing both models, it was concluded that in order to predict percentage cost overturn of construction project CBR model is less capable in prediction. Beside this, for testing the effectiveness of CBR model with weight assignment method, best results were obtained when applied absolute standardized coefficient (β). Better results were given by Feature counting method than the original value of (β). Thus, to predict cost overrun percentage for construction projects best approach for industry practitioners was provided. Moreover, it was also provided methodology to build regression and CBR models for cost overrun percentage for researchers. For future work It is suggested to use computer implementation for case-based reasoning mode for easily implementation.

REFERENCES

- [1] Oberlender, G.D., and Trost, S.M. "Predicting Accuracy of Early Cost Estimates Based on Estimate Quality." J. of Constr. Eng. and Manage., 127, 173-182, 2001.
- [2] Kim, K. J., and Kim, K." Preliminary Cost Estimation Model Using Case-Based Reasoning and Genetic Algorithms" J. Comput. Civ. Eng., 24 (6), 499-505, 2010.
- [3] Koo, C.W., Hong, T., Hyun, C., and Koo, K. "A CBR-Based Hybrid Model for Predicting a

- Construction Duration and Cost Based on Project Characteristics in Multi-Family Housing Projects" *Can. J. Civ. Eng.*, 37, 739-752, 2010.
- [4] Kangari, R. "Risk Management Perceptions and Trends of U.S. Construction." *J. of constr. Eng. and Manage.*, 121(4), 422-429, 1995.
- [5] Ibbs, W.C., and Ashley, D.B., "Impact of Various Construction Contract Clauses." *J. of Constr. Eng. and Manage.*, 113(3), 501-521, 1987.
- [6] Sonmez, R., Ergin, A., and Birgonul. T. "Quantitative Methodology for Determination of Cost Contingency in International Projects", *J. of Manage. in Eng.*, 23,(1), 35-39, 2007.
- [7] Trost, S., M., and Oberlender, G., D. "Predicting Accuracy of Early Cost Estimates Using Factor Analysis and multivariate Regression", *J. of constr. Eng. and Manage.*, 129 (2), 198-204, 2003.
- [8] Abu Hammad, A. A., Ali, S. M. A., Sweis, G., J., and Basher, A." Prediction Model for Construction Cost and Duration in Jordan", *Jordan J. of Civil Engineering*, 2(3), 250-266, 2008.
- [9] Lowe, D.J., Emsley, M.W., and Harding, A. "Predicting Construction Cost Using Multiple Regression Techniques" *J. of Constr. Eng. And Manage.*, 132(7), 750-758, 2006.
- [10] Phaobunjong, K. "Parametric Cost Estimating Model for Conceptual Cost Estimating of Building Construction Projects", Ph.D. thesis, University of Texas, Austin, TX, 2002.
- [11] Attala, M., and Hegazy, T., "Predicting Cost Deviation in Reconstruction Projects: Artificial Neural Networks Versus Regression" *J. of Constr. Eng. and Manage.*, 129 (4), 405-411, 2003.
- [12] Dogan, S.Z, Arditi, D., and Gunaydin, H.M. "Determining Attribute Weights in A CBR Model for Early Cost Prediction of Structural System" *J. of Constr. Eng. and Manage.*, 132(10), 1092-1098, 2006.
- [13] Hegazy, T., and Ayed "Neural Network Model for Parametric Cost Estimation of Highway Projects" *J. of Constr. Eng. and Manage.*, 124 (3), 210-218, 1998.
- [14] Nassar, K.M., Gunnarsson, H.G., and Hegab, M.Y. "Using Weibull Analysis for Evaluation of Cost and Schedule Performance" *J. of Constr. Eng. and Manage.*, 131 (12), 1257-1262, 2005.
- [15] Ji, S.H., Park, M., and Lee, H.S." Cost Estimation Model for Building Projects Using Case-Based Reasoning" *Can. J. Civ. Eng.*, 38, 570-581, 2011.
- [16] Ryu, H. "Construction Planning Methodology Using Case-Based Reasoning (COPLA-CBR)" Ph.D., thesis, Seoul, National Univ., Seoul. Korea, 2007.
- [17] Duverlie, P., and Castelain, J.M., "Cost Estimation during Design Step: Parametric Method Versus Case Based Reasoning Method" *Adv. Manuf. Technol.* 15 (12), 1999.
- [18] Kholif, W., Hosny, H., and Sanad, A. "Analysis of Time and Cost Overruns in Educational Building Projects in Egypt" *Int. J. of Eng. and Technical Research (IJETR)*, ISSN: 2321-0869, 10(1), 2013.
- [19] Aziz, R., F., "Factors Causing Cost Variation for Constructing Wastewater Projects in Egypt" *Alexandria Eng. J.*, 52, 51-66, 2013.
- [20] Pal, S.K., and Shiu, S.C.K. "Foundations of Soft Case-Based Reasoning" Wiley, Hoboken, N.J., 2004
- [21] Kaming, P. F., Olomolaiye, P. Holt, G., and Harris F. "Factors Influencing Construction Time and Cost Overruns on High-Rise Projects in Indonesia, 1997.
- [22] Miaoulis, G. and Michener, R.D. "An Introduction to Sampling". Dubuque, Iowa: Kendall/Hunt Publishing Company, 1976.
- [23] Israel, G.D." Determining Sample Size" Agricultural Education and Communication Department, Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, University of Florida, 1992.
- [24] Cochran, W.G. "Sampling Techniques", 2nd Ed., New York: John Wiley and Sons, Inc, 1963.
- [25] Alreck, P. L., and Settle, R. B. "The survey research handbook." Richard D. Irwin, Inc., Homewood, Ill, 1985.
- [26] Oppenheim, A. N. "Questionnaire Design, Interviewing, and Attitude Measurement". Pinter publisher, London, 1992