Mansoura Engineering Journal

Volume 37 | Issue 1 Article 10

10-10-2021

Modeling Cost Overrun in Construction Projects Case Based Reasoning Versus Regression Analysis.

A. El-Kholy
Associate Professor., Civil Engineering Department., Delta Higher Institute of Engineering and Technology, Mansoura, Egypt.

Recommended Citation
El-Kholy, A. (2021) "Modeling Cost Overrun in Construction Projects Case Based Reasoning Versus Regression Analysis.," *Mansoura Engineering Journal*: Vol. 37 : Iss. 1 , Article 10.
Available at: https://doi.org/10.21608/bfemu.2021.201523

MODELING COST OVERRUN IN CONSTRUCTION PROJECTS: CASE-BASED REASONING VERSUS REGRESSION ANALYSIS

"نمذجة الزياده فى تكلفة مشروعات التشييد: مقارنه بين تقنية الاستدلال اعتمادا على الحالات السابقه و تحليل الانحدار"

A. M. EL-KHOLY[1]

الملخص العربي

هذا البحث يقدم نموذجين للتنبؤ بنسبة زيادة التكلفة فى مشروعات التشييد و ذلك للمقاولين بمصر. النموذج الأول يعتمد على تحليل الانحدار. تم تجميع ٤٤ سبب من الأسباب التى تؤثر على زيادة تكلفة تلك المشروعات وذلك من الدراسات السابقة. تم إجراء استبيان على مقاولي التشييد لتقييم الأهمية النسبية لهذه الأسباب من وجهة نظر المقاولين. تم الحصول على ١١ سبب هى الأكثر تأثيرا على نسبة زيادة التكلفة و هي تمثل المتغيرات المستقلة في النموذج المقترح. تم تجميع بيانات تخص حدوث أو عدم حدوث الأسباب التي تؤدي لزيادة التكلفة فى مشروعات التشييد والنسبة المناظرة لزيادة التكلفه التي حدثت (المتغير التابع) لتلك المشروعات وذلك لعدد ٣٠ مشروع. تم تقسيم تلك المشروعات لجزئين: الجزء الأول يشمل ٢٠ مشروع لبناء النموذج وقد أظهرت النتائج وجود علاقة خطيه قويه بين المتغير التابع والأسباب السابقة. هذه الأسباب هى: الوضع المالي للمالك، المسار النقدي للمقاول، طريقه اختيار المقاولين (مناقصه عامه أم مناقصه محدوده)، زيادة تكلفه المواد بسبب التضخم، طبيعة المنافسه فى مرحلة العطاء، تقلبات العملة، حجم المشروع (صغير أم كبير)، التأخير في التصميم والموافقات، المخاطر من المالك بسبب اختلاف الكميات، الرسومات (مفصله أم لا)، تقدير غير دقيق للمواد. الجزء الثاني يشمل ١٠ مشروعات وذلك لاختبار هذا النموذج. النموذج الثاني يعتمد على تقنية الاستدلال اعتماداً على الحالات السابقة ،**Case Based Reasoning**، و هى وسيلة فعالة لاستخدام المعارف المكتسبة من التجارب السابقة لتقديرنسبة الزياده فى التكلفه في أعمال التشييد. تم اختبار قابلية التطبيق للنموذجين حيث تم التنبؤ بنسبة التأخير للجزء الثاني من المشروعات . وقد أظهرت النتائج أن النموذج المعتمد على تحليل الأنحدار يعطى نتائج أدق فى التنبؤ عن النموذج المعتمد على تقنية الاستدلال اعتماداً على الحالات السابقة. من ناحية أخرى، وجد أن تطبيق القيمة المطلقه للمعامل موحد القياس (**Standardized Coefficient(** β **)**) كطريقة لوزن (**attributes**) يزيد من دقة التنبؤ لنسبة الزياده فى التكلفه يليه تطبيق طريقة الوزن المتساوى (**Feature Counting**) ثم القيمة الأصلية للمعامل موحد القياس (β).

ABSTRACT

This paper presents two models for predicting cost overrun percentage in construction projects for the contractors in Egypt. The first model based on regression analysis. 44 factors that impact cost performance in construction projects gathered from literature. A questionnaire survey was made on construction contractors in Egypt to evaluate the relative importance of these causes from contractors' perspective. Eleven factors were obtained as the most significant causes that lead to cost overrun and these are the independent variables of the proposed model. Data was collected for occurrence of the previous factors on yes/no basis and the corresponding cost overrun percentage (dependent variable) for 30 construction projects and was divided into two sets. The first set contains 20 projects for model building. The results revealed that there was a strong linear relationship between cost overrun percentage and the previous 11 causes that significantly affect cost overrun of projects. These causes are: financial condition of the owner, cash flow of contractor, method of procurement (open tender or selective tender), material cost increase due to inflation, competition at tender stage (aggressive or not), fluctuations in the currency that the payment will be made, project size (small or large), delay in design and approval, risk retained by client for quantity variations, drawings (detailed or not), and inaccurate material estimating. The second set contains 10 projects for validation purposes. The second model is a case based reasoning (CBR) model. CBR method can be an effective means of utilizing knowledge gained from past experience to estimate percentage cost overrun in construction works. Validation of the two models using projects of the second set revealed that regression model has prediction capabilities higher than that of CBR model. Applying the absolute value of standardized coefficient (β) as attribute weight method provides the highest prediction accuracy of cost overrun percentage. Also, feature counting gives results better than the original value of (β).

[1] Assistant Professor, Civil Eng. Dept., Delta Higher Institute of Eng. and Tech., Mansoura, Egypt

Accepted March 31, 2012

Keywords: Regression Analysis; Questionnaire Survey; Construction Projects;
Case Based Reasoning Model

1. INTRODUCTION

The accuracy of early cost estimates in engineering and construction projects is extremely important to both owners and project teams (Oberlender and Trost) [1]. Decision making in the early stage of a project has a significant impact on the project. To evaluate alternatives, quick and accurate decision making is needed under a limited definition of scope and constraints in available information and time (Kim and Kim) [2]. However, limited and uncertain information on the project and a complex correlation among various factors that affect the project's construction cost, makes it difficult to predict and manage pertinent task (Koo et al.)[3].

Several studies have attempted to determine the factors creating risk for construction projects. Kangari [4] has conducted a survey to study the risk attitudes of large U.S. construction firms. Among the 23 risk factors included in this survey, labor, equipment and material availability, labor and equipment productivity, defective design, changes in work, differing site conditions, safety, delayed payment on the contract, and quality of work were presented as risks with high importance. Ibbs and Ashly [5], focused on the contract related factors which play an important role in the allocation of risks between the owner and the contractor. Sonmez et al. [6], found that country risk rating, material availability, type of contract, advance payment were the major factors impacting contingency decisions of the contractors. Trost and oberlender [7] developed a multivariate regression model to predict cost estimate accuracy for capital projects.

The previous studies used various methodologies to solve the problem of predicting construction cost, cost contingency, and cost overrun for construction projects. Some of the methods used in the previous studies include:

- Statistical methods such as multiple regression analysis (MRA) for predicting construction cost (Abu Hammad et al.) [8], (Lowe et al.) [9], and (Phaobunjong) [10]. Attalla and Hegazy [11] presented a regression model for predicting cost overrun of reconstruction projects. Sonmez et al. [6] and Trost and Oberlender [7] presented models for predicting cost contingency.
- Repetitive learning methods such as artificial neural networks (ANN) for predicting construction cost (Dogan et al.) [12] and (Hegazy and Ayed) [13]. Attala and Hegazy [11] presented an ANN model for predicting cost overrun of reconstruction projects in addition to the regression analysis mentioned above.
- Stochastic methods such as Monte-Carlo simulation (MCS). Nassar et al. [14] conducted a simulation model for predicting the construction cost.
- Analogical methods such as Case-based reasoning (CBR) for predicting the construction cost (Ji et al.) [15], (Kim and Kim) [2], (Ryu) [16], (Dogan et al.) [12] (Duverlie and Castelin) [17].

Koo et al. [3] stated that such methodologies have distinct characteristics in terms of applied fields, analysis of data, methods of system establishment, and types of results. Multiple regression analysis arrives at the result through statistical analysis, but its result is too linear to be used as a standardized model. Artificial neural networks is more accurate than MRA, but it has a black box that cannot explain the structure of the model. Monte carlo simulation has the function of analyzing the outlier using the probability approach (Koo et al.) [3]. On the other hand, CBR has characteristics that are similar to humans' heuristic approach in which decisions are made on experience.

In this paper a MRA model is developed for predicting cost overrun percentage for construction projects. A second model is developed depending on CBR for the purpose of comparison.

The major objectives of this paper are as follows: (1) Investigate the causes that significantly affect cost overrun of construction projects in Egypt; (2) Propose two models: based on regression analysis and case based reasoning method to predict percentage of cost overrun for construction projects for the contractors in Egypt.

2. RESEARCH SCOPE AND METHODOLOGY

In the current research two proposed predictive models are intended to be applicable for predicting cost overrun percentage of construction projects with contract value LE 2.5 millions or more. The contractors awarded these contracts represent the second and first classes as classified by Egyptian Federation for Construction & Building Contractors. These models are based on regression analysis and case based reasoning. A standard methodology will be adopted. As an initial step to meet the objectives, previous research papers that deal with causes of cost overrun in construction projects were reviewed in previous section to investigate causes of cost overrun in these projects. CBR technique is explained. A list of causes of cost overrun in construction projects is prepared to collect data about the significance of these causes through questionnaire survey. The next step is to analyze the survey results to obtain the most significant causes of cost overrun to be incorporated into the predictive models. Building regression based model is then demonstrated and a numerical example is prepared to show how the model predicts cost overrun percentage of a project. The next step is to apply the case based model to an example project to show how the model performs step by step. The last step of this research is to validate the proposed models. Based on the results of validation, the prediction accuracy of the two models is compared and conclusions are drawn.

3. CASE BASED REASONING

CBR is the process of retrieving previous cases similar to a new problem, solving the new problem by adapting previously determined solutions of similar previous cases, and storing the new successful solution for future use (Pal and Shiu)[18]. Ji et al.[15] stated that CBR utilizes knowledge gained from past experiences and can be viewed as an effective method for estimation in construction. It has been observed that CBR methods can increase the accuracy of construction cost estimates (Karshenas and TSE ; Chua and Loh ; Yi ; An et al.) [19-22]. Kim and Kim [2] reported that CBR requires usually four steps; case representation, case retrieval, case adaptation, and case retaining. Cases are represented by attributes describing the circumstance of the problem and its solution. Similar previous cases best matching the new problem are retrieved. The solution(s) of the retrieved cases are adapted to fit the new problem. New solution(s) are retained for future use once it has been approved. Ji et al. [15] explained that there are two challenges related to the retrieval process that still needs to be addressed. One issue is the computation of attribute similarity which is particularly important during the retrieval process. For calculating attribute similarity: if an attribute is of nominal scale, and its value in a previous case is the same as in a new case, then the attribute is rated as one, otherwise it is rated zero (Koo et al.) [3]. On the other hand, if an attribute is either of interval scale or ratio scale, it is scored by Eq. 1. The second challenge is how to assign the attribute weight values that enable the most similar case to be identified by an index of corresponding features. Koo et al. [3] declared that more than one methodology can be used for calculating attribute weight, as follows:

- Feature counting: this method applies the same weight to all the attributes.
- MRA: this method uses the original standardized value of the coefficient (β) as the attribute weight. On the other hand, Kim et al. [23] used the absolute value of standardized coefficient (β).
- ANN: this method uses the sensitivity coefficient as the attribute weight.

In the current research, the attributes are of nominal scale, since the data gathered depends on a yes/no basis. Thus, any attribute will be rated one if its value in case base is the same as in test case, otherwise it will be rated zero. Also, for calculating the attributes weights: feature counting, original value of the standardized value (β), and its absolute value will be used for the purpose of comparison to improve the prediction capacity of CBR model.

To calculate the case similarity, the attribute similarity is multiplied by its weight of importance and summed up to obtain total similarity score of each case as given in Eq. 2 (Kim and Kim) [2].

$$F_{AS} = \begin{cases} \dfrac{Min\,(Av_{testcase}, Av_{retrievedc\ ase})}{Max\,(Av_{testcase}, Av_{retrievedc\ ase})} & \text{If}\, F_{AS} \geq \text{MCAS} \\ 0 & \text{If}\, F_{AS} < \text{MCAS} \end{cases} \tag{1}$$

Where F_{AS} is the function of the attribute similarity, $Av_{testcase}$ is the attribute value of test case, $Av_{retrievedcase}$ is the attribute value of the retrieved case, and MCAS is the minimum criterion for scoring the attribute similarity.

$$S_i = \sum_{j}^{n} I_{ij} *W_j \tag{2}$$

Where i =case identification number; j =attribute identification number; S_i= similarity value of case i ; I_{ij}=similarity value of attribute j ; and W_j= weight of attribute j .

4.FACTORS THAT IMPACT COST PERFORMANCE IN CONSTRUCTION PROJECTS

In this study, 44 factors are identified as causes of cost overrun in construction projects were gathered from literature: Oberlender and Trost [1]; Sonmez et al [6].; Attala and Hegazy [11] ; Burroughs and Juntima [24]; Iyer and Jha [25]; Touran [26]; Dissanayaka and Kumaras a [27]; Dissanayaka and Kumaras b [28]; Bacon and Besant [29]; Kaming et al. [30]; Akinsola et al.[31] as shown in Table 1. These factors serve as the independent variables in the predictive model of cost overrun percentage for construction projects.

5. QUESTIONNAIRE SURVEY

A questionnaire was developed to collect data about the significance of the factors that impact cost of construction projects compiled in Table 1. The questionnaire was divided into two main parts. The first part gathered basic information about the experience of the respondent, experience of the company, and volume of work of the company. In the second part 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 priority scaling for occurrence frequency 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. The participants were asked to assign a number from 1 to 5 to each factor for both occurrence frequency and severity according to its significance. Besides, the questionnaire 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 (see appendix A). In other words, if in a past project, one of the previous factors occurred, the respondent assigns yes to this factor otherwise, he assigns no. Also, the actual percentages of cost overrun for these projects are gathered.

To determine the sample size of the questionnaire three criteria usually will need to be specified : the level of precision, the level of confidence or risk, and the degree of variability in the attributes being measured (Miaoulis and Michener) [32]. Israel [33] reported that, the level of precision is the range in which the true value of the population is estimated to be. This range is often expressed in percentage points, (e.g., ±10 percent). Thus, if a researcher finds that 60% of respondents in the sample have adopted a recommended practice with a precision rate of ±10%, then he or she can conclude that between 50% and 70% of respondents in the population have adopted the practice. For the confidence or risk level, if a 95% confidence level is selected, then 95 out of 100 samples will have the true population value within the range of precision specified earlier. The third criterion, the degree of variability in the attributes being measured refers to the distribution of attributes in the population. The more heterogeneous a population, the larger the sample size required to obtain a given level of precision. A proportion of 50% indicates a greater level of variability than either 20% or 80%. This is because 20% and 80% indicate that a large majority do not or do, respectively, have the attribute of interest. Because a proportion of 0.5 indicates the maximum variability in a population, it is often used in determining a more conservative sample size.

For population that are large Cochran [34] developed Eq. 3 to yield a representative sample size for large population (n_0). If the population is small, one can use Eq. 4., where(n) is the sample size for small population.

$$n_0 = \frac{Z^2 pq}{e^2} \tag{3}$$

$$n = \frac{n_0}{1 + \frac{(n_0 - 1)}{N}} \tag{4}$$

Where Z^2 is the abscissa of the normal curve that cuts off an area at the tails (1 – equals the desired confidence level, e.g., 90%), e is the desired level of precision, p is the estimated proportion of an attribute that is presented in the population, and q is $1-p$. The value for Z is found in statistical tables which contain the area under the normal curve. N is the population.

In the current research, the population is 465, which represents the number of contractors works in construction projects with LE 2.5 millions or more, this number was obtained from Egyptian Federation for Construction & Building Contractors. The population is large, thus Eq. 3 is applied first for determining an initial sample size (n_0). A confidence level, 90% is assumed, thus Z=1.65 from normality tables, p is assumed 0.5, e is assumed (±15%). Substituting about: Z, p, q, and e in Eq. 3, results in an initial sample size

n_0=30.25. Substituting about: n_0 and N in Eq. 4, results in sample size n=28.5. Logically the anticipated response rate, will not be 100%. Accordingly, the questionnaire was sent to 43 contracting companies specialized in construction projects. Some of the questionnaires were sent via mail after contacting the participants through telephones, whereas, the other part was through individual meetings. Most of the participants were at the level of general managers.

6. SURVEY RESULTS AND ANALYSIS

A total of 30 questionnaires were completed and returned. The response rate was 69.8 %. This response rate is considered acceptable for a survey focusing on gaining responses from industry practitioners (Alreek and Settle) [35]. The respondents included general managers, technical office managers, and construction managers. All the participants are involved in building projects in addition to other specializations. 82% of them are involved in public water and sewage projects. Also, 42% of them are involved in civil works (bridges, roads, and airports). 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 (Oppenhiem) [36].

To give additional credibility for the findings of this survey, the participants were asked about their length of experience and length of experience of their companies. 89% of the respondents have an experience more than 10 years, whereas, 57% have an experience more than 20 years. 92% of the companies have an experience more than 10 years, whereas 50% have an experience equal to or greater than 25 years. 78% of the companies have an annual volume of work more than LE 25 millions, whereas 42% have an annual volume of work equal to or more than LE 250 millions.

In order to assess the significance of the identified factors, an importance index for each factor was calculated, as illustrated in Eq. (5), by multiplying the frequency of occurrence by the degree of severity or impact. Frequency occurrence refers to the probability that any factor given in Table 1 occurs in a project and contributes to its cost overrun. Whereas, degree of severity refers to the negative impact that the factor contribute to the project cost overrun. The importance indices were used to measure the relative weight for each factor. The relative importance weight (RIW) was computed using Eq. 6. The factor financial condition of the owner, for example, if its assigned (4=often), for frequency of occurrence, this means that the interviewer assigns a 80% probability for the occurrence of this factor effect in previous projects according to his experience. In these projects this factor contributed to these projects cost overrun. On the other hand, if this factor assigned (4=severe) for the degree of severity, this means that the impact of this factor was severe on these projects' cost overrun. As an another example is the factor "site layout", if its assigned (3=usually), for frequency of occurrence, this means that the interviewer assigns a 60% probability for the occurrence of this factor effect in previous projects according to his experience. In these projects this factor contributed to these projects cost overrun. On the other hand, if this factor assigned (2=little effect) for the degree of severity, this means that the impact of this factor was little on these projects' cost overrun. Table 1 shows the factors arranged in descending order according to their corresponding RIW, such that the factor received the highest RIW is assigned rank equal to one.

$$\text{Importance Index (II)} = \text{Occurrence frequency*degree of severity} \qquad (5)$$

$$\text{Relative Importance Weight (RIW)} = \frac{\Sigma \ \prod \text{*corresponding no. of respondents}}{\text{Total no. of respodents}} \qquad (6)$$

Table 1: Factors that impact cost 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

Financial condition of the owner comes out as the most important factor contributing to cost overrun in construction projects, it was ranked the first. Cash flow of contractor received the second rank. It seems that, if the contractor suffers from negative cash flow for most or all other projects, he fails to finance the project under consideration, thus the project is extended, which leads to cost overrun. The third ranked factor was the method of procurement (open tender or selective tender). It seems that, when open tender applied as a method for project procurement to the contractors, they decrease cost contingency in projects or it is neglected completely. Material cost increase 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 cost overrun for the construction project. On the other hand, Kaming et al. [30] found that this factor is among three main causes of cost overrun. Competition at tender stage (aggressive or not) received the fifth rank (see Table 1), it seems that when the competition is aggressive, contractors decrease cost contingency or it is neglected completely. Fluctuations in the currency that the payment will be made ranked 6. Project size (small or large) ranked 7. It seems that cost overrun appears to be more predominant among smaller projects compared to larger ones. Delay in design and approval, Risk retained by client for quantity variations, drawings (detailed or not), and inaccurate material estimating were ranked: 8, 9, 10, and 11, respectively. Causes received RIW less than 10 will not be considered in the predictive model 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.

7. REGRESSION BASED MODEL

Data for 30 construction projects was collected. These data include the occurrence of factors presented in Table 1 on a yes/ no basis, and the corresponding actual cost overrun percentage as presented in the appendix. The data was divided into two sets. The first set contains 20 projects for the purpose of model building. The second set contains 10 projects for validation purposes. An initial experimentation with a regression model that includes all 11 variables using SPSS 13 software was performed. Forward-stepping and backward-stepping methods were used. Forward stepping begins with entering the most significant variable at the first step, and continues adding and deleting variables until none can significantly improve the fit. Backward stepping, on the other hand begins with all candidate variables then removes the least significant variable at the first step and continues until no insignificant variable remains. Forward- stepping or backward-stepping technique gave the same model for predicting the percentage of cost overrun for construction projects depending on 11 variables (see Table 3) with a squared multiple R= 0.83. This indicates that the model is able to explain 83 % of the variability in the data, which is an excellent indicator of the model's expected performance. The underlying formula of the model is as follows:

$$\begin{aligned}&\text{Percentage cost overrun}\\&= 0.214+ 0.046\ (\text{Financial condition of the owner})\\&+ 0.201\ (\text{Cash flow of contractor}) + 0.345\ (\text{Method of procurement (Open tender or}\\&\text{Selective tender)})- 0.177(\text{Material cost increase due to inflation}) -0.197\ (\text{Competition at}\\&\text{tender stage (aggressive or not)})-0.108\ (\text{Fluctuations in the currency that the payment}\\&\text{will be made})-0.078(\text{Project size (small or large)})-0.284\ (\text{Delay in design and approval})\\&+0.08\ (\text{Risk retained by client for quantity variations}) + 0.184\ (\text{Drawings (detailed or}\\&\text{not)})+0.08\ (\text{Inaccurate material estimating}).\end{aligned} \quad (7)$$

Each of the 11 variables can have a 0 (unused), or 1 (used) value.

To show how the model predicts the cost overrun percentage, it can be used in predicting cost overrun percentage for a project with the following characteristics:

Financial condition of the owner was bad (1); cash flow of contractor was bad (1); method of procurement was open tender (1); material cost increased due to inflation was occurred (1); competition at tender stage was not aggressive (0); fluctuations in the currency that the payment will be made were occurred (1); project size was small (1); delay in design and approvals was occurred (1); there was a risk retained by the client for quantity variations (1); drawings was detailed (0); material estimating was accurate (0)

The predicted cost overrun percentage will be obtained as follows:

Cost overrun percentage= 0.214 +0.046* 1 +0.201*1+0.345*1-0.177*1-0.197*0-0.108*1-0.078*1-0.284*1+0.08*1 +0.184*0 +0.08*0 = 0.239.

This result means that the predicted cost overrun percentage is 23.9 %. This model will be validated later using the second set of projects.

8. CASE-BASED REASONING MODEL

Based on previous cases (first set of projects), a case base is developed. Then, those cases that are similar to the new cases are retrieved from the case base in order to estimate cost overrun percentage of the new cases. To retrieve similar cases, the similarity values are calculated by multiplying each similarity value $(I_{11}, I_{12}, I_{13}, ..., I_{1j}, ..., I_{m1}, I_{m2}, I_{m3}, ..., I_{mj})$ of each attribute (factor) for a case in the case base and new case by corresponding attribute weight ($W_1, W_2, W_3, ..., W_j$) and then summing all of them. The weights of attributes are variables. The case with the highest similarity value is used to estimate cost overrun percentage of the new case. In the following subsections, these processes are described in details.

8.1 Case Representation and Attributes

Each case is represented by the attributes identification and dependent variable (percentage of cost overrun). In attributes identification, the attributes are presented which, are the previous 11 factors that impact cost of construction projects included in the regression model. The attributes are used in calculating degree of similarity between a new case (test case) and each case in the case base. In current research, all the attributes are of nominal scale, since each attribute assigned a value of one if occurred, otherwise zero. Thus, a similarity value of attribute is assigned one if its value in each case of case base is the same as its value in test case, otherwise, 0. Also, three methods are used for calculating the attributes weights: feature counting, standardized coefficient (β) of MRA, and the absolute value of the standardized coefficient (β) for the purpose of comparison to improve the prediction capacity. It must be noted that, standardized coefficient (β) is a regression coefficient in a standard format as given in the results of SPSS software. Also, if this coefficient used as a method of calculating attributes weights, in the case based reasoning model, MRA must run first. Feature counting method is used for calculating attributes weights without running MRA. The similarity value of each attribute is multiplied by its weight resulted from any method of the three previous methods and summed to obtain the similarity value of each case (see Eq. 2).

8.2 Matching and Retrieval

In comparison with the new cases in which percentage cost overrun will be estimated, the most similar case (the case with the highest similarity value from a case base) to the new case is retrieved. If more than one case in case base has the same similarity value, the author suggests using the mean value of cost overrun percentage for these cases.

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	Coeff.
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	

8.3 Adaptation

In this study, all the attributes are of nominal scale, thus no adaptation is used to adjust cost overrun percentage for the new cases.

8.4 Case Retaining

In this research, the new cases are retained for future use, i.e in finding the solution of the second new case, the first new case is used among the cases of the case base. Also, in finding the solution of the third new case, the first and second cases are used among the cases of the case base and so on.

9. EXAMPLE PROJECT

To show how the model performs, an example project is solved step by step to predict cost overrun percentage for this project. The data for this project was obtained from Arab Contractor company works in Egypt. It is the construction of an hotel in Ismalia city, Egypt, with a lump sum contract. The contract value was LE 6 millions and the duration was 3 years. The actual cost overrun for this project was 25%. The project characteristics were as follows: Financial condition of the owner was bad (1); cash flow of contractor was bad (1); method of procurement was open tender (1); material cost increased due to inflation was occurred (1); competition at tender stage was not aggressive (0); fluctuations in the currency that the payment will be made were occurred (1); project size was small (1); delay in design and approvals was occurred (1); there was a risk retained by the client for quantity variations (1); drawings was detailed (0); material estimating was accurate (0). These values are given in Table 4. Such that if the factor has been occurred it is assigned a value of 1, otherwise its assigned zero. This project is the previous project used in predicting cost overrun in the regression based model.

Calculating the attributes similarity

Table 4 shows the data for actual 20 construction projects obtained from the questionnaire (the first set of projects). These data are the attributes (factors) values according to their occurrence and the corresponding actual percentage of cost overrun. The attribute F_1 is assigned a similarity value one as its value in the example project is similar to that of case 1. On the other hand, F_5 (for example) in case 1 is assigned a value of one, whereas its assigned a value of zero in the example project, accordingly its similarity value is zero. The similarity value of all other atrributes in case 1, are given in Table 4. Other cases are calculated as presented in example project.

Calculating attributes' weights

Table 5 shows, the previously mentioned three methods for calculating attributes weights. In feature counting method each attribute receive a weight of (1/11 =0.0909). In the second and third methods, the standardized coefficient (β) and its absolute value resulted from regression model are used as weights to the attributes (see Table 5).

Similarity value of each case

Applying Eq. 2, the total similarity value for each case in case base and test case can be calculated by multiplying similarity value of each attribute by its weight. The total similarity value (S_1) between case 1 and test case (example project) using the second method (standardized coefficient (β)) for weighting attributes (for example), is calculated as follows:

$$S_1=1*0.127+1*0.582+1*0.998-*1*0.513-0*0.596-1*0.328-1*0.238-1*0.69+0*0.232+0*0.398+0*0.238=-0.062$$

Calculating similarity value for all cases in case base and test case using [standardized coefficient (β)] for weighting attributes, (for example), revealed that case 3 received the highest value (1.656). Thus, case 3 is retrieved and the predicted value of cost overrun percentage for the example project is 15% (actual value for case 3).

Case reataining

The new case, which consists of attributes of example project and 15% cost overrun percentage is retained for future use in addition to cases of case base.

10. MODELS VALIDATION

A comparison between the regression model estimate and the case based reasoning model estimate is shown in Table 6. It provides the actual cost overrun percentage, predicted cost overrun percentage, and the analysis of the average percent error for 10 projects including the example project (these are projects of the second set). In case based reasoning model, three methods for calculating attributes weights were used as presented previously. In general, the regression based model shows prediction accuracy better than that of case based reasoning model. Average % error=34.8 for regression based model, whereas this percentage is varied for CBR model according to weight assignment method for attributes. Best results for CBR model are obtained when applying absolute standardized coefficient (β) as assignment method for attributes

Table 4: Profile of cases for case base and test case (example project)

Case base No.	Attributes											% actual cost overrun
	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	
1	1	1	1	1	1	1	1	1	0	1	1	0.20
2	1	1	1	1	1	1	1	1	0	1	1	0.25
3	1	1	0	1	0	1	1	1	0	1	1	0.15
4	1	1	1	1	0	1	1	1	1	1	0	0.55
5	0	0	0	0	1	0	0	0	0	0	0	0.05
6	1	1	1	1	1	1	1	1	0	1	0	0.15
7	1	1	1	1	1	1	1	1	1	1	1	0.40
8	0	0	1	1	1	0	0	1	1	1	0	0.15
9	1	1	1	0	0	0	1	1	0	1	1	0.75
10	1	1	1	1	1	1	1	1	1	1	1	0.15
11	1	1	1	1	1	1	1	1	1	1	1	0.20
12	0	0	0	0	0	0	0	1	0	1	0	0.05
13	1	0	1	0	1	0	0	1	0	1	0	0.35
14	0	0	0	0	0	0	0	0	0	1	0	0.35
15	1	1	0	1	0	0	0	0	0	0	0	0.25
16	0	0	0	1	1	0	0	0	1	1	0	0.15
17	1	1	1	1	0	1	1	1	0	1	0	0.25
18	0	1	1	1	1	1	0	1	0	1	1	0.30
19	1	1	0	0	0	0	0	1	0	1	0	0.35
20	1	0	1	0	0	0	0	1	1	0	0	0.40
Example Project	1	1	1	1	0	1	1	1	1	0	0	0.25
Attr. Sim. for case1	1	1	1	1	0	1	1	1	0	0	0	-

Table 5: Attribute weights by methods

No.	Attribute	Method of weighting		
		Features counting	Standarized coefficient (β)	Absolute standarized coefficient (β)
F1	Financial condition of the owner	0.0909	+0.127	0.127
F2	Cash flow of contractor	0.0909	+0.582	0.582
F3	Method of procurement (open tender or selective tender)	0.0909	+0.998	0.998
F4	Material cost increase due to inflation	0.0909	-0.513	0.513
F5	Competition at tender stage (aggressive or not)	0.0909	-0.596	0.596
F6	Fluctuations in the currency that the payment will be made	0.0909	-0.328	0.328
F7	Project size (small or large)	0.0909	-0.238	0.238
F8	Delay in design and approval	0.0909	-0.690	0.690
F9	Risk retained by client for quantity variations	0.0909	+0.232	0.232
F10	Drawings (detailed or not)	0.0909	+0.398	0.398
F11	Inaccurate material estimating	0.0909	+0.238	0.238

(average % error= 40.7). This percentage is 45.9 for feature counting method.. On the other hand, this percentage is 58.2 when applying the original value of standardized coefficient (β).

11. CONCLUSIONS AND FUTURE RECOMMENDATIONS

This paper investigated the effect of causes of cost overrun affecting construction projects in Egypt through a questionnaire survey. These causes were established from literature. The questionnaire survey used a structured format to obtain information related to the occurrence of the previous causes in actual projects on a yes/no basis. Based on the results of the questionnaires a relative importance weight was established for each cause to quantify its effect on project cost performance. It was intended that causes received a relative importance weight higher than 10 are significant and incorporated into the model as independent variables. Accordingly, 11 significant causes were identified. The dependent variable was the cost overrun percentage.

Two models were developed to predict cost overrun percentage in construction projects. The first model based on regression analysis. Data of 20 projects was used for model building, while the data of remaining 10 projects was used for validation purposes. The best model was found accurate in predicting cost overrun percentage contains the previous 11 causes. These are: financial condition of the owner, cash flow of contractor, method of procurement (open tender or selective tender), material cost increase due to inflation, Competition at tender stage (aggressive or not), fluctuations in the currency that the payment will be made, project size (small or large), project size (small or large), delay in design and approval, risk retained by client for quantity variations, drawings (detailed or not), and inaccurate material estimating.

The second model based on case based reasoning. Validation of the two models revealed that regression model has prediction capabilities higher than that of CBR model in predicting cost overrun percentage for construction projects. On the other hand, testing the case based reasoning model's effectiveness with respect to the weight assignment method for attributes, revealed that best results are obtained when applying absolute standardized coefficient (β). Feature counting method gave results better than the original value of (β). This research provides an approach for industry practitioners to predict cost overrun percentage for construction projects. On the other hand, it provides researchers with a methodology to build regression and case based reasoning models for cost overrun percentage prediction. Computer implementation for case based reasoning model is suggested for future research, for easily implementation.

Table 6: Models Validation

Case Project	Percent cost overrun output				
	Project Actual	Regression Model	Case based reasoning Model		
			Feature counting	Standardized coefficient (β)	Absolute standardized coefficient (β)
Example Problem	25	23.9	55	15	55
2	40	22.6	22.5	75	23.3
3	30	22.6	22.5	75	23.3
4	15	9.8	25	35	25
5	10	15.8	15	25	15
6	45	30.6	25	75	25
7	25	27.3	20	15	25
8	20	30.6	25	75	25
9	45	34	40	75	35
10	60	49.5	42.5	35	40
Average % Error	0.00	34.8	45.9	58.2	40.7
Average % Error $= \left|X_{actual} - X_{estimated}\right| / X_{estimated} \times 100$					

12. 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. Pal, S.K., and Shiu, S.C.K." Foundations of Soft Case-Based Reasoning" Wiley, Hoboken, N.J., 2004.
19. Karshenas, S., and Tse, J. "A Case –Based Reasoning Approach to Construction Cost Estimating" .J. Comput. in Civil Eng., 113-123, 2002.
20. Chua, D.K.H, and Loh, P.K." CB-Contract: Case Based Reasoning Approach to Construction Contract Strategy Formulation" "J. Comput. in Civil Eng., 20 (5), 339-350, 2006.
21. Yi, J. "A Study on Case- Based Forecasting Model for Monthly Expenditures of Residential Building Project" Korean J. of Constr. Eng. and Manage., 79 (1), 128-137, 2006.
22. An, S. H., Kim G., and Kang, K." A Case Based Reasoning Cost Estimating Model Using Experience by Analytic Hierarchy Process" Building and Environment, 42 (7), 2573-2579, 2007.
23. Kim, G., Kim, S.,and Kang, K. "Comparing Accuracy of Prediction Cost Estimation Using Case-Based Reasoning and Neural Networks" J. Architectural Institute of Korea, 20, (5), 93-102, 2004.
24. Burroughs and Juntima " Exploring Techniques for Contingency Setting" AACE Transactions, 2004.
25. Iyer, K.C., and Jha, K.N." Factors Affecting Cost Performance: Evidence from Indian Construction Projects" Intern. J. of Project Management, 23: 283-295, 2005.
26. Touran, A." Probabilistic Model for Cost Contingency" J. of Constr. Eng. And Manage., 129 (3), 280-284, 2003.
27. Dissanayaka, S.M, and Kumaraswamy, M.M "Comparing Contributors to Time and Cost Performance in Building Projects, Building and Environment, 34, 31-42, 1999a.
28. Dissanayaka, S.M, and Kumaraswamy, M.M "Evaluation of Factors Affecting Time and Cost Performance in Hong Kong Building Projects" Engineering, Construction

and Architectural Management , 6(3), 287-298,1999b.

29. Bacon, R.R. and Besant, J "Estimating Construction Costs and Schedules: Experience with Power Generation Projects in Developing Countries, Energy Policy, 26(4), 317-333, 1998.
30. Kaming, P. F., Olomolaiye, P. Holt, G., and Harris F. " Factors Influencing Construction Time and Cost Overruns on High-Rise Projects in Indinesia,1997.
31. Akinsola, A.O., Potts, K.F., Ndekugri, I., and Harris F.C." Identification and evaluation of Factors Influencing Variations on Building Projects" Intern. J. of Project Management, 15(4), 263-267, 2005.
32. Miaoulis, G. and Michener,R.D. "An Introduction to Sampling". Dubuque, Iowa: Kendall/Hunt Publishing Company, 1976.
33. 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.
34. Cochran, W.G. "Sampling Techniques", 2nd Ed., New York: John Wiley and Sons, Inc, 1963.
35. Alreck, P. L., and Settle, R. B. "The survey research handbook.'' Richard D. Irwin, Inc., Homewood, Ill, 1985.
36. Oppenheim, A. N. "Questionnaire Design, Interviewing, and Attitude Measurement". Pinter publisher, London, 1992.

Appendix. Questionnaire on causes of cost overrun affecting construction projects in Egypt

Part "A"

1. **Company Name:**
2. **Respondent's Title or Position:**
3. **Respondent's Experience in Construction Field:**
   a. Less than 5 years.
   b. From 5 years up to less than 10 years.
   c. From 10 years up to less than 15 years.
   d. From 15 years up to less than 20 years.
   e. From 20 years up to less than 25 years.
   f. 25 years or more.
4. **Type of Work:**
   a. Residential Buildings.
   b. Administrative and commercial Buildings.
   c. Civil Works (Bridges- Highways-airports).
   d. Civil Works (Water and Sewage projects).
5. **Company Experience in Construction Field:**
   a. Less than 5 years.
   b. From 5 years up to less than 10 years.
   c. From 10 years up to less than 15 years.
   d. From 15 years up to less than 20 years.
   e. From 20 years up to less than 25 years.
   f. 25 years or more.
6. **Annual Construction Volume**
   a. Less than 10 millions.
   b. From 10 to less than 15 millions.
   c. From 15 to less than 25 millions.
   d. From 25 to less than 40 millions.
   e. From 40 to less than 60 millions.
   f. From 60 to less than 80 millions.
   g. From 80 to less than 100 millions.
   h. From 100 to less than 250 millions.
   i. More than 250.

Part "B"

7. Please mark against the suitable rate for the following causes that affect cost overrun percentage of construction projects for frequency of occurrence and degree of severity. For frequency of occurrence use the priority scale: 5=always, 4=often, 3=usually, 2=sometimes, 1= rarely. For degree of severity use the priority scale:5=very sever, 4=sever, 3= somewhat sever, 2=little effect, 1= very little effect. Please fill column no. 5 for the following table on yes/ no. bases using data of an actual construction project.

Table A: Frequency of occurrence and degree of severity for causes of cost overrun in construction projects

NO. (1)	Cause of cost overrun (2)	frequency of occurrence (3)					degree of severity (4)					cause occurred (5)	
		5	4	3	2	1	5	4	3	2	1	yes	No.
1	Project size (small or large).												
2	Adequacy of schedule requirements.												
3	Adequacy of quality requirements.												
4	Adequacy of safety and environmental requirements.												
5	Contract type (unit price or lump sum)												
6	Adequacy of dispute settlement procedure												
7	Advanced payment amount												
8	Drawings (detailed or not)												
9	Delay in design and approvals												
10	Complexity of the Design												
11	Similar project experience												
12	Time allowed for preparation of estimate												
13	workload												
14	Financial condition of the owner												
15	Fluctuations in the currency that the payment will be made												
16	Unknown geological conditions												
17	Weather conditions												
18	Site access												
19	Site congestion												
20	Cash flow of contractor												
21	Prequalification of contractors												
22	Site layout												
23	Liquidated damages												
24	Inspection and testing												
25	Scope change by owner												
26	Design change												
27	Time variance												
28	Quality standards and specifications												
29	Client characteristics												
30	Risk retained by client for quantity variations												
31	Level of construction complexity related to new technology and payment modality												
32	Estimated cost												
33	How the estimate is prepared?												
34	What is known about the project?												
35	Location of project												
36	Equipment percentage												
37	Conflict among project participants												
38	Ignorance and lack of knowledge												
39	Reluctance in timely decision												
40	Competition at tender stage (aggressive or not)												
41	Inaccurate material estimating												
42	Material cost increase due to inflation												
43	Difference between low bid and owner's estimate												
44	Method of procurement (open tender or selective)												

8. According cost overrun occurred in the actual project given in Table A, mark the actual cost overrun percentage in Table B for the project characteristics given in column 5 of Table A.

Table B: Actual cost overrun percentage for the project given in table A

Percentage	%5	%10	%15	%20	%25	%30	%35	%40	%45	%50
Mark										
Percentage	%55	%60	%65	%70	%75	%80	%85	%90	%95	%100
Mark										