Forecasting the Final Cost of Iraqi Public School Projects Using Regression Analysis

Dr. Zeyad S. M. Khaled
College Engineering, Alnahrian University/ Baghdad.
Dr. Qais Jawad Frayyeh
Building and Construction EngineeringDep, University of Technology/ Baghdad.
Gafel Kareem Aswed
Building and Construction EngineeringDep, University of Technology/ Baghdad.
Email:gafelkareem@gmail.com

Received on:29/5/2014 & Accepted on:8/1/2015

ABSTRACT

The actual final cost of public school building projects, like other construction projects, is unknown to the owner till the final account statement is prepared. An attempt to predict the final cost ofsuch projects before work starts, using backward elimination regression analysis technique is carried out.The study covers two story (12 classes) school projectsawardedby the lowest bid system. Records of (65) school projects completed during (2007-2012) are employed to develop and verify the regression model. Based on experts'convictions, nine factors are considered to have the most significant impact on the final cost.Hence they are used as model inputparameters. These factorsare;awarded bid price, average bid price, estimated cost, contractor rank, resident engineerexperience,project location, number of bidders, year of contracting, and contractual project duration. It was found that the developed regression model have the ability to predict the final cost (FC) for school projects, as an output, with avery good accuracy havinga correlation coefficient(R)of(93%),determination coefficient (R^2)of(86.5%)and average accuracy percentage of(92.02%).

Keywords: cost estimation, cost modeling, regression analysis, and school project.

التنبؤ بالكلفة النهائية لمشاريع المدارس العراقية الحكومية بطريقة تحليل الانحدار

الخلاصة

تبقى الكلفة النهائية الحقيقية لمشاريع المدارس،كغيرها من المشاريع الإنشائية، غير معروفة لرب العملحتى يتماعداد الحساباتالختامية.تم في هذا البحث استخدام تقنيةتحليل الانحداربطريقة الحذفلأجل التنبؤ بالكلفة النهائية لمشاريع المدارس قبل بدء العمل بالمشروع. وقد شملت الدراسة تحليل بيانات (65) مشروع مدرسةبطابقين، ذات (12) صف، محالة بنظام الإحالة لأوطأ العطاءات. جميعها مكتملة البناء ومستلمة استلاماً نهائياً من قبل رب العملخلال السنوات (2007-2012). وقد تم تحديد تسع عوامل اساسية مؤثرة في الكلفة النهائية لمثل هذهالمشاريع، وذلك من خلال استبيان أراء ذوي الخبرة والاختصاص،وتم استخدامها كمدخلات في بناء

https://doi.org/10.30684/etj.33.2A.18
2412-0758/University of Technology-Iraq, Baghdad, Iraq

النموذج الرياضي للتنبؤ بالكلفة النهائية.وهذه العوامل هي: سعر العطاء الفائز,ومعدل اسعار العطاءات,والكلفة التخمينية,ودرجة تصنيف المقاول,وخبرة المهندس المقيم,وموقع المشروع,وعدد المناقصين,وسنة التعاقد,ومدة التنفيذ التعاقدية.ولقد تبين من الدراسةأنالنموذج المعتمدلها القدرةعلى التنبؤبالكلفةالنهائيةلمشاريع المدارس، كمخرج للنموذج،بدرجةعاليةمن الدقة وبمعامل ارتباط يساوي(93٪)،ومعامل تحديد يساوي(86.5٪)، ومتوسط نسبة دقة يساوي(92.02٪).

INTRODUCTION

Construction projects costsare influenced byseveralfactors. These factorsare related to project characteristics, construction teamsand market conditions. When unexpected events occurduring the execution phase of construction projects,their final costs aredriven up. Most of such events are uncontrollable factors that increase the gap between thecontract award priceand the final completion cost. It is greatly important that the client should know what contingencies he must have in hand to ensure his project final completion in time. Lack of information about these factors, lack of relevant data, and weak expectations of possible circumstances to be faced by the project are the main challenges facing researchers in thisessence.This research attempts to use real measurable parameters, to be in hand before the project starts, as predictors for the expected final cost of school projects.

Research Objectives

This research aims at the following objectives:

- To explore the factors that can be used to predict the final cost of school building projects before starting works.
- To raisethe efficiency of estimating initial costs using data already in hand.
- To build a mathematical model using multiple regression analysis to predict cost deviation in school building projects before starting works.

Research Hypothesis

At the project start phase, it can be said that awarded bid price, average bid price, estimated cost, contractor rank, resident engineerexperience,project location, number of bidders, year of contracting, and contractual project durationare good predictors to the final cost of public school building projects before starting works.

Research Justification

The reasons forcarrying out this research are:

- The large number of under construction school projects accompanied with everlasting cost overrun and the ever growing demand on additional school buildings in Iraq.
- The need of knowingan accurate anticipated final cost of a construction projects before starting works, is highly essential in budgeting concerns, especially in contingency allocation.

Suitabilty Of Multiple Regression

The objective of most parametric costs estimating approaches is to use some historical cost data and try to find a functional relationship between changes in cost and the factors affecting the final cost.The regression technique is a statistical modeling method that can be used for analysis and prediction in different knowledge domains. Multiple regressionestimation models are well established and widelyused in cost estimation. They are effective due to their well-definedmathematical procedure, as well as being able to explain the significance of eachvariable and the relationships between them. Basically, regressionmodels are intended to find the linear combination of variableswhich bestcorrelates with dependent variables.The general regressionequation is expressed asfollows**[1]**:

$$Y = A_o + A_1 I_1 + A_2 I_2 + \ldots . A_n I_n \quad (1)$$

Where

Y is the total estimated final result, A_o is aconstant estimated by regression analysis, A_1, A_2, … A_n are coefficients also estimated by regression analysis, given the availability of some relevant dataI_1, I_2,…I_n as measured distinguishable variables that may help in estimating Y **[2]**.

Literature Review

Literature review shows a variety of ways used to predict the project final cost and deviations. Many variables were used as predictors in those studies.

Williams **[3]** concluded five mathematical models to predict the final cost of highway construction using low bids in five states in USA as independent variable. From competitive bidding of highway construction projects, the low bid price and the cost of the completed contract were obtained for each project. These models aimed at predicting the projectfinal cost according to the low bid as the only input.

Wibowo and Wuryanti**[4]**studied education building projects in Indonesia to prepare early stage cost estimate models. They found that the estimated cost at earlier stages could be predicted according to the total project area.

Olatunji**[5]** collected data of (137) public contract projects executed between (2003) and (2007) in Nigeria. Lowest/winner bid, average bid, consultant's cost estimate, gross floor areawere the model variables to predictthe final construction cost. The conducted regression model has an adjusted $\mathbf{R^2}$ value of (0.949).

Mahamid and Amund **[6]** investigated the statistical relationship between actual and estimated cost of road construction projects. Data collected from (169) road construction projects awarded in the West Bank in Palestine over the years (2004–2008) were analyzed. The study concluded that (100%) of road projects in Palestine suffer from cost deviation. theydeveloped a regression model with a coefficient of determination (R^2)of (0.96).

Bedford**[7]** studied the risks of excessive competition in the Canadian public sector that award contracts solely to the low bid. It has been concluded that the bidding process is a good indicator to the final cost and possible cost escalations.

Ganiyu and Zubairu **[8]** developed a predictive cost model for public building projects in Nigeria using principal components regression. The study showed that the project cost basicallydepends on factors related to; adequacy of equipment, experience in similar projects, time allowed for bidding, level of technology, client commitment to time, repetitive work,design complexity, communications,project scope, construction complexity, and previous relationship with the client.

Mohd et al.**[9]**studied the historical data of (83) school projects in the Malaysian public sector. Multiple regression analysis was used to predict the effect of the lowest bid, average tender price, and the winning tender in the interpretation of the deviation in cost estimation.The regression model from mean bids showed that the project size, number of tenders, type of schools, and location are the best-fitted predictors to explain biased estimates.

Aziz**[10]** investigated and ranked factors perceived to affect cost variation in the Egyptian wastewater projects.It was discovered that factors such as: lowest bid procurement method, additional works, bureaucracy in bidding, tendering method, wrong method of cost estimation, and funding problems were crucialin causing cost variation, while, inaccurate cost estimation, mode of payments, unexpected ground conditions, inflation, and pricefluctuation are less important.

Datacollaction

The initial parameters that are intended to be used in the model were collected from the literature review of previous studies as shown in Table (1). A questionnaire form has been directed to (50) local experts in order to determine the most significant factors in predicting the final cost of school projects before it starts. Fifty questionnaires were directed to owner's representative engineers in the public sector. Thirty two respondents forms, forming (64%) of the total number of questionnaires, have successfully been submitted. The respondents were asked to select the parameters that they believe is most important in developing the mathematical model. Likert’s scale of five ordinal measures of importance (from 1 to 5) is used. An efforthad been made tolimit the number of variables that affects the cost model. Nine out of eleven variables were identified and analyzed as independent variables of the regression equation based on the respondents' opinions. These variables are: ($\mathbf{I_1}$) accepted bid price, ($\mathbf{I_2}$) average bid price, ($\mathbf{I_3}$) estimated cost, ($\mathbf{I_4}$) contractor rank, ($\mathbf{I_5}$) experience of the resident engineer (R.E), ($\mathbf{I_6}$) location of project, ($\mathbf{I_7}$) number of bidders, ($\mathbf{I_8}$) year of contracting, and ($\mathbf{I_9}$) project duration. To build the regression model a historical data is collected from (68) completed schools projects in Karbala province. These projects are executed during the years (2007-2012). By eliminating the incomplete records, the number of projects put under consideration for the final selected model became (65) projects. The projects were awarded under the lowest bid tendering system having the same design and number of classrooms.

Table (1): Influential factors affecting the final cost of school projects

InfluentialFactors		**Supporting Previous Studies**
$\mathbf{I_1}$	Accepted bid price	Elhag [11],Williams[3], Olatunji[5]
$\mathbf{I_2}$	Average bid price	Olatunji[5], (Mohd et al.[9]
$\mathbf{I_3}$	Estimated cost	Olatunji [5], Mahamid and Amund[6]
$\mathbf{I_4}$	Contractor rank	Ganiyu and Zubairu[8]
$\mathbf{I_5}$	Experience of R. E.	Ganiyu and Zubairu[8]
$\mathbf{I_6}$	Project location	Mohd et al.[9],
$\mathbf{I_7}$	Number of bidders	Mohd et al. [9]
$\mathbf{I_8}$	Year of contracting	Al-zwainy [12]
$\mathbf{I_9}$	Contractor duration	Elhag [11]
$\mathbf{I_{11}}$	Owner duration	Ayman et al. [13]
$\mathbf{I_{12}}$	Second lowest bid	Bordat et al. [14]

Modelformulation

Previous studies showed different methods used to study the relation between the final cost and factors believed to influence that final cost. In this research a back elimination regression technique isadopted toanalyze historical cost data in order to provide a powerful modelto assist budgeting and cost estimating before work starts. The Statistical Package for Social Science SPSS and MS Excel are used to develop a suitable model.In order to remove lineartrend from the data, transformationsby taking the natural logs of some of thevariablesareapplied.Then a simple linear model is developed using in each run, the natural log of thefinal project cost (FC) as the dependent variable,the natural logs of; accepted bid price ($\mathbf{I_1}$), average bid price ($\mathbf{I_2}$), estimated cost ($\mathbf{I_3}$), and the untransformed parameters of; contractor rank ($\mathbf{I_4}$), experience of R.E. ($\mathbf{I_5}$), location of project ($\mathbf{I_6}$), number of bidders ($\mathbf{I_7}$), year of contracting ($\mathbf{I_8}$), contractor duration ($\mathbf{I_9}$) as independent variables.

SPSS (version 20) is used for data analysis. Backward elimination technique is adopted to develop the regression model as in Tables(2) and (3). The procedure of this technique is to enter all nine variables in the model equation first, then sequentially remove the variable with the smallest partial correlation with the dependent variable in each run.

Table (2): Summary of Analysis Results

FC-Model	R	R^2	Adjusted R^2	Standard Error	Residual Mean Square
1	0.930a	0.865	0.841	0.1232519	0.015
2	0.930b	0.865	0.844	0.1220738	0.015
3	0.930c	0.865	0.847	0.1209313	0.015
4	**0.930d**	**0.865**	**0.850**	**0.1198686**	**0.014**
5	0.929e	0.862	0.850	0.1199525	0.014
6	0.927f	0.860	0.850	0.1199546	0.014
7	0.926g	0.857	0.849	0.1201290	0.014

Where

a. Predictors: (Constant), I_9, I_4, I_6, I_7, I_5, I_2, I_8, I_3, I_1

b. Predictors: (Constant), I_9, I_4, I_6, I_7, I_5, I_8, I_3, I_1

c. Predictors: (Constant), I_9, I_4, I_7, I_5, I_8, I_3, I_1

d. Predictors: (Constant), I_9, I_4, I_7, I_5, I_3, I_1(the best model)

e. Predictors: (Constant), I_4, I_7, I_5, I_3, I_1

f. Predictors: (Constant), I_4, I_7, I_3, I_1

g. Predictors: (Constant), I_7, I_3, I_1

h.Dependent Variable: FC

Table (3): Model Coefficients

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.(*p*-Value)
	B	Std. Error	Beta		
(Constant)	**2.032**	1.235		1.646	**0.106**
I_1	**0.344**	0.147	0.358	2.343	**0.023**
I_3	**0.569**	0.147	0.560	3.860	**0.00031**
I_4	**-0.025**	0.020	-0.067	-1.283	**0.205**
I_5	**0.004**	0.004	0.059	1.111	**0.272**
I_7	**-0.028**	0.013	-0.109	-2.096	**0.041**
I_9	**0.00022**	0.000	0.065	1.037	**0.304**

Resulted Equation

After applying multipleregressionanalysis on the historical data of thewhole (60)school projects, theresultedfinal construction cost estimation equation is:

$$FC= 2.032 + 0.344\mathbf{I_1}+ 0.569\mathbf{I_3}- 0.023\mathbf{I_4}+ 0.004\mathbf{I_5}- 0.028\mathbf{I_7}+ 0.00022\mathbf{I_9} \quad \text{…. (2)}$$

This model is chosen based on the smallest Standard Error of Estimate which is (0.1198686) and the Residual Mean Squarewhich is(0.014).

According to Tabachnick and Fidell**[1]**advice, the relative importance of the independent variables is assessed by examining their respective standardized coefficients i.e. Beta values in Table (3). Predictors with higher standardized coefficients such as: ln accepted bid price ($\mathbf{I_1}$), ln estimated cost ($\mathbf{I_3}$) and number of bidders ($\mathbf{I_7}$) are more important to the regression equation than those with lower valuessuch as contractor rank ($\mathbf{I_4}$), experience of R.E. ($\mathbf{I_5}$) and contractor duration ($\mathbf{I_9}$).Therefore values of $\mathbf{I_1}$, $\mathbf{I_3}$, and$\mathbf{I_7}$indicate a highly significant regression fit. It can be concluded that $\mathbf{I_1}$,$\mathbf{I_3}$, and $\mathbf{I_7}$contribute significantly to the regression model. The small constants of $\mathbf{I_5}$ and $\mathbf{I_9}$in the model equation refer to the small effect of experience of R.E and the contractor duration. The exclusion of the average bid price ($\mathbf{I_2}$), location ($\mathbf{I_6}$) and the year of contracting ($\mathbf{I_8}$) parametersis because of their insignificance.

Multi-Colliinearity Assessment

To assess multi-collinearity among the variables, tolerances and variance inflation factors (VIF) are examined as shown in Table (4). Tolerance refers to the proportion of the variance of that variable that is not accounted for by other predictors in the modeland is calculated using the formula ($1–R^2$) for each variable. The range of tolerances is from (0)i.e. perfect collinearity, to (1)i.e. no collinearity. A tolerance with values less than (0.1) typically indicates a multi-collinearity problem. Variance inflation factor (VIF) is another index for the diagnostic of multi-collinearitywhichis just the inverse of the tolerance value. The high value of (VIF) for a variable indicates that there is a strong association between that variable and other remaining predictors [**1**].Variables that have high tolerances will definitely have small variance inflation factors. A variance inflation factor in excess of (10) indicates a multi-collinearity problem **[15]**. Since the final cost model predictors havetolerances and (VIF) values that does not violates the aforementioned criteria, therefore, multi-collinearity is not a serious problem in this analysis.

Table (4): Collinearity test

Model	VIF	Tolerance
$\mathbf{I_1}$	9.180	.109
$\mathbf{I_3}$	8.276	.121
$\mathbf{I_4}$	1.070	.934
$\mathbf{I_5}$	1.091	.916
$\mathbf{I_7}$	1.069	.936
$\mathbf{I_9}$	1.538	.650

Model Validation

One of the most important steps in developing a cost model is to test its accuracy and validity. This process is also refers to as the model validation.It involves testing and evaluating the developed model with some test or validation data. The validation data should be some representative data from the targeted population but haven't been used in the development of the model. In this study, the validation data is extracted from the same historical data filebut forfiverandomly selectedadditional projects. They are not a part of the (60) projects used in the development of the model. The predicted cost of these five projectscomputed using the model equation are compared withreal cost records and the results of this comparisonis shown in Table (5). It is evident now that the model performed very well.Its predictions deviate only by (–8.702%) to (6.513%).

Table (5): Comparison of predicted and observed final costs

FC(observed) in IQD	FC (predicted) in IQD	*Deviation %
688482184.5	733325333.7	6.513335
1158046071	1057263708	-8.70279
832546804.6	760502532.7	-8.65348
1852866989	1877430643	1.325711
1594777396	1473269621	-7.61911

*Deviation %= {(Predicted cost – Observed cost)/Observed cost*100}**[16]**.

To assess the validity of the derived equation of the model for the final cost of school project (FC), the natural logarithm (Ln) of predicted values of (FC) are plotted against the natural logarithm (Ln) of real values for validation data set as shown in Figure (1).

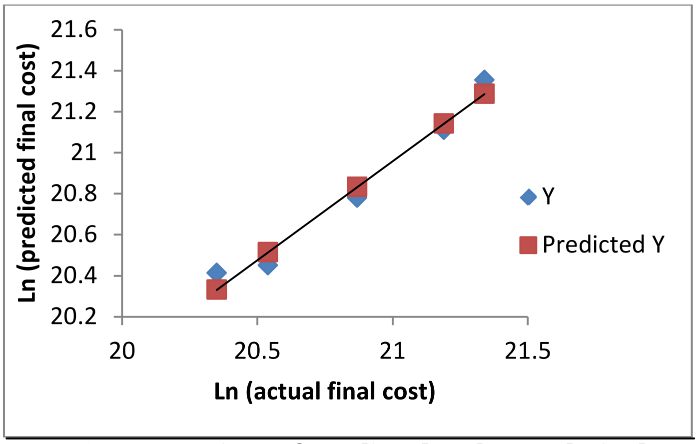

Figure (1): Comparison of Predicted and Actual Final Costs

The coefficient of determination (R^2) is found to be (97%), therefore it can be concluded that this model shows a good agreement with the actual measurements. It is finally clear that this model for school building projects in Iraq has the generalization capability for any data set used within the range of data used in the development of the (MR) model.

Model Evaluation

The statistical measures used to measure the performance of themodels included**[17]**:

i. Mean Percentage Error (MPE)which is one of the most important measures of accuracy of a proposed model. It is the mean of the absolute percentage differences

ii. between the predicted and the actual values:

$$MPE = \left\{\sum_{j=1}^{n}\left[\frac{A-E}{A}\right]/n\right\} * 100 \qquad \ldots(3)$$

Where:

A = actual value

E = estimated value (predicted value)

n = total number of cases (**5** for validation)

iii. Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{\sum_{j=1}^{n}(E-A)^2}{n}} \qquad \ldots(4)$$

iv. Mean Absolute Percentage Error (MAPE),

$$\text{MAPE} = \frac{\left\{\sum_{j=1}^{n}\frac{|A-E|}{A} * 100\right\}}{n} \qquad \ldots(5)$$

Average accuracy percentage (AA %):

AA% = 100% -MAPE …(6)

v. The Coefficient of Determination (R2)

vi. The Coefficient of Correlation (R) is a measure that is used to determine the relative correlation and the goodness-of-fit between the predicted and observed data and show how well the model outputsmatch the target value.

These statisticalmeasures of the regression Model (FC) in equation (2) are presented in Table (6).

Table (6):Statistical Measures Results For Regression model

Description	Statistical parameters
MPE	3.98%
RMSE	0.83
MAPE	7.97%
AA%	92.02%
R^2	86.5%
R	93%

The(MAPE) and (AA) generated by (MR) model of (FC) are found to be (7.97%) and (92.02%) respectively.The (R^2) value is (86.5%) which indicates that the most variability in the total cost is explained by the terms in the model. Therefore,it can beconcluded that this (MR) model of (FC)showsgood agreement with theactual measurements.

Conclusions

1. Backward elimination techniques is used to develop the regression model on the historical data of the school projects, resulting in the final construction cost perdition equation with (R^2 =86.5%).
2. The most significant model parameters are: accepted bid price ($\mathbf{I_1}$), estimated cost ($\mathbf{I_3}$), Contractor rank ($\mathbf{I_4}$), resident engineer experiencein years($\mathbf{I_5}$), number of bidders ($\mathbf{I_7}$), and contractor duration ($\mathbf{I_9}$).
3. The developed model shows agood agreement with theactual measurements based on the accuracy measurements.
4. It can easily calculate the expected cost deviation which is the difference between contractual sums and predicted final costs obtained from the developed regression model.
5. Other future work estimating parameters like cost per unit area, unit volume, classroom, or pupil can be calculated.

Recommendations

1. Data for another design of schools can be studied to further confirm the relationship between the independent parameters and final cost.
2. The developed model can be checked for applicability on other typical type of school projects such as (14), (16) and (18) classes.
3. Attention must be given to the documentation and feedback of data in order to achieve efficient and effective updated information and to audit a deferent accounting and financial procedures.

REFERANCES

[1].Tabachnick, Barbara G. and Fidell, Linda S., “Using Multivariate Statistics”, fifth Edition, 2007,Pearson Education Inc.

[2].Stevens, I. James P., “Applied multivariate statistics for the social sciences” 2009, 5th edition- Taylor & Francis Group, LLC.

[3].Williams, Trefor P., “Predicting Final Cost for Competitively Bid Construction Projects Using Regression Models”, International Journal of Project Management, Vol. 21, 2003, Pp. 593–599.

[4].Wibowo, A. and Wuryanti, W., “Capacity Factor Based Cost Models for Buildings of Various Functions”, Civil Engineering Dimension, Vol. 9, No. 2, 2007, Thomson Gale TM, Pp. 70–76.

[5].Olatunji, O. A., “A Comparative Analysis of Tender Sums and Final Costs of Public Construction and Supply Projects in Nigeria”, Journal of Financial Management of Property and Construction, 13(1), 2008, Emerald Group Publishing Limited, Pp. 60-79.

[6].Mahamid, I and Amund, B “Cost deviation in road construction projects: the case of Palestine”, Australasian Journal of Construction Economics and Building, 12 (1), 2012, Pp. 58-71.

[7].Bedford, Thomas, “Analysis of the Low-Bid Award System in Public Sector Construction Procurement”, MSc. thesis in Civil Engineering, 2009, University of Toronto.

[8].Ganiyu, B. O. and Zubairu, I. K., “Project Cost Prediction Model Using Principal Component Regression for Public Building Projects in Nigeria”, Journal of Building Performance, Vol. 1 Issue 1, 2010, Pp. 21-28.

[9].Mohd Azrai Azman, Zulkiflee, A. S., Suraya Ismail, “The Accuracy of Preliminary Cost Estimates in Public Works Department (PWD) of Peninsular

Malaysia", International Journal of Project Management Vol. 31, 2013, Elsevier, Pp. 994-1005.

[10].Aziz, Remon Fayek, "Factors Causing Cost Variation for Constructing Waste Water Projects in Egypt", Alexandria Eng.Journal Vol.52, 2013, p. 51–66.

[11].Elhag, Taha Mahmoud Salih, "Tender Price Modeling: Artificial Neural Networks and Regression Techniques", PhD Dissertation, 2002, University of Liverpool.

[12].Al-Zwainy, Fai'q Mohammed Sarhan, "The Use of Artificial Neural Network for Estimating Total Cost for Highway Construction Projects", Ph.D. Thesis, 2008, College of Engineering, University of Baghdad.

[13].Ayman, A. Abu Hammad, Souma M. Alhaj Ali, Ghaleb J. Sweis and Adnan Bashir, "Prediction Model for Construction Cost and Duration in Jordan", Jordan Journal of Civil Engineering, Volume 2, No. 3, 2008,Pp. 250-266.

[14].Bordat, C., B. G. McCullouch, S. Labi, and K. C. Sinha, "An Analysis of Cost Overruns and Time Delays of INDOT Projects", Publication FHWA/IN/JTRP-2004/07.Joint Transportation Research Program, Indiana Department of Transportation and Purdue University, West Lafayette, Indiana,.

[15].Pallant, Julie, "SPSS Survival Manual a step by step guide to data analysis using SPSS", 4th edition, 2011, Allen & Unwin, Australia.

[16].Memon, A. H., "Structural Modeling of Cost Overrun Factors in Construction Industry", PhD dissertation, faculty of Civil and Environmental Engineering, University Tun Hussein Onn Malaysia, 2013.

[17].Sulewska, Maria J., "Applying Artificial Neural Networks for analysis of geotechnical problems", Computer Assisted Mechanics and Engineering Sciences, 18, 2011, Pp.231–241.