Indian Journal of Science and Technology, Vol 12(7), DOI: 10.17485/ijst/2019/v12i7/138685, February 2019
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Developing a Prediction Model to Predict the Construction Project Cost by Using Multiple Linear Regression Technique

Abbas Mahde Abd[1*], Nidal Adnan Jasim[2] and Fatima Saleh Naseef [2]

[1]Department of Architectural Engineering, College of Engineering, University of Diyala, Iraq; abbas.mahde@gmail.com
[2]Department of Civil Engineering, College of Engineering, University of Diyala, Iraq; nidaladnan100@gmail.com, fatimasaleh5420@gmail.com

Abstract

Objectives: Prediction cost of construction project requires large information and data about the project. This makes the prediction cost very complex at the early stage because of limitation of data and information at this stage. The aim of the study is building prediction model to predict cost of construction project in Iraq. **Method:** To develop the prediction model, Multiple Linear Regression technique (MLR) with Weighted Least Square (WLS) was used. The researcher use 501 set of historical cost data gathered in Iraq for period (2005-2015) for developing the model. The cost of twenty five items of project are used for cost forecasting by MLR model and they involved cost of (excavation the foundation works, Landfill works, filling with sub-base works, Construction works under moisture proof layer, Construction works above moisture proof layer, Construction works of sections, ordinary concrete for walkways, reinforced concrete foundation, reinforced concrete column, reinforced concrete lintel, reinforced concrete slabs, reinforced concrete beams, reinforced concrete stair, reinforced concrete for the sun bumper, plaster finishing works, cement finishing works, Plastic Paints, Pentellite paints, Stone packaging, Works of placing marble, Ceramic works for floor, Ceramic works for walls, Flattening (two opposite layers of lime), Flattening (Tiling). **Findings:** The result shows that MLR with WLS has the capability to predict construction cost with a height coefficient of correlation 95.8%, degree of accuracy 98.97% and smallest mean absolute percentage error 1.03%. **Applications:** MLR with WLS have shown to be a promising method for using in the initial stage of construction projects when only limited data and incomplete information set is preparing for cost analysis.

Keywords: Cost Estimation, Construction Projects, Regression, MLR Model, WLS

1. Introduction

Construction manager consider cost is very important factor. If the construction project matching the budget with the cost, schedule on time and quality determined by the employer then the project is considered success[1]. When using a weak strategy or inappropriate budget or schedule forecasting the project may be failure[2]. Therefore, the success of any construction project depends on the accurate estimation cost[3]. Accordingly, estimation cost in initial phase has important role in any construction project[4]. Because of importance of estimation cost in early phase and limited data during the early stage of construction, construction managers leverage their experience, knowledge and estimators to estimate cost of project. As such, intuition plays an important role in decision-making[2].

Many researches were attempted to predict the cost of construction[5]. In[6] developed models by using multiple linear regression techniques to estimate initial cost of road projects features such as soil conditions, soil drill ability and terrain conditions. In[7] develop models for

*Author for correspondence

estimating the construction cost of buildings by using Multiple Regression Techniques. The researcher use 286 sets of historical data gathered from the United Kingdom. They establish 41 independent variables, the researcher uses five important variables such as gross internal floor area (GIFA duration, function, piling and mechanical installations in all six models. In[8] develop models by utilities MLR to estimate productivity of marble finishing work by using 100 set of historical data gathered in Iraq for various kinds of construction projects. The researcher concludes that MLR have the capability to estimate the productivity for finishing works with height degree of accuracy. A recent study[9] used MLR to develop model for estimating the cost of communication projects in Iraq. The data that used to develop model were 45 construction projects. They used several significant independent variables that act on communication for developing MLR model. From the result, the researchers found regression analysis techniques proven its ability to prediction cost with very height of accuracy.

The scope of the study is using multiple linear Regression technique to developing and assesses a prediction model that can used to predict the final cost of projects, through the procedure below:

- Determination input and output for model.
- Using the MLR to develop a mathematical prediction model to predict the budget of project.
- Cheek the developed model by making the verification and validation through calculate the degree of accuracy of model and the coefficient of correlation between real cost and prediction cost.

2. Methodology of Study

Methodology of this study divides to theoretical work and field work:

- Theoretical part: This includes the review of literatures that associated to cost estimation concept and multiple linear regression analysis and its utilization in prediction the budget of construction projects.
- Field part: This includes data collection, data analysis and choosing of inputs, then building the prediction Model. And validation this Model.

3. Application of MLR

M any reference clarified the concept of MLR technique. The research submitted by 8 explain this technique in an easy way. Multiple linear regression considers strong powerful technique that can use as prediction tool and help the engineers and researchers to know the correlation between input and output variables. It can tentatively frame[10]:

$$Yi = a+ a1\ xi1 + a2\ xi2 + \ldots\ldots + ei \quad (1)$$

where:

i=1,……,n

- xi1 and xi2 are independent value
- yi the desired output.
- ei the error components are supposed to be in normal variables with mean zero and variance σ2. and a0, a1, ….,ap are unknown regression coefficient.

4. Weighted Least Square Regression

WLS has not specific type of equation to know the correlation between input and output variables, contrasting the linear and nonlinear regression that is associated to an equation. This type of regression used with equation that linear or nonlinear in parameters. It is working by weights that connected with each observation, into the fitting criterion. Weight value refers to the accuracy of information contained in the associated observation. Improving the weighted fitting criterion to obtain the parameter estimates permits the weights to define the contribution of each observation to the last parameter estimates. It is significant to observe that the weight for each observation is given relative to the weights for another observation; so different sets of absolute weights can have identical special effects[11]. The researcher use WLS to determine the regression coefficient.

5. Input and Output for Model

Model input (independent variable) variables for cost estimation model were consisting of twenty five variables are cost of excavation the foundation works, Landfill works, filling with sub-base works, Construction works under moisture proof layer, Construction works above moisture

proof layer, Construction works of sections, ordinary concrete for walkways, reinforced footing, reinforced columns, reinforcement lintel, reinforcement slabs, reinforcement beams, reinforcement stair, reinforced concrete for the sun bumper, plaster finishing works, cement finishing works, Plastic Paints, Pentellite paints, Stone packaging, Works of placing marble, Ceramic works for floor, Ceramic works for walls, Flattening (two opposite layers of lime), Flattening (Tiling) Symbolizes by (x1,x2,.....x25) Sequentially. Output of model is total cost of project.

6. Development MLR

The software employed to develop the MLR model is Statistical Package for the Social Sciences version 24. Table 1 shows the statistical analysis of data. Backward elimination method is implemented to develop the regression model. The procedure of this method is to enter all input variables in the model equation and then gradually excludes. The input that has least partial correlation with the output variable is considered first for removal as shown in Table 2.

The standardized coefficient such as Beta value in Table 2 refers to comparative importance of inputs variables. The input variables with higher standardized coefficients such as Flattening (Tiling) (x25) is more effective on the regression equation than those with lower standardized coefficients such as construction works of sections reinforcement (x6) and concrete stairs (x13)[12]. The researcher concludes from the table that all indepen-

Table 1. Descriptive statistics

variables	N	Range	Minimum	Maximum	Std. Deviation
x1	501	33972430	596370	34568800	8219596.048
x2	501	51248000	1552000	52800000	9567490.102
x3	501	29392000	1498900	30890900	7141612.462
x4	501	1.8E+08	4701600	1.85E+08	41549178.41
x5	501	6.63E+08	27926000	6.91E+08	113465324.4
x6	501	58880000	1120000	60000000	11158927.06
x7	501	76930000	1620000	78550000	12647415.85
x8	501	5.23E+08	10403000	5.34E+08	101741597.4
x9	501	1.31E+08	1408700	1.32E+08	26653439.84
x10	501	3.57E+08	2568500	3.6E+08	81512835.77
x11	501	7.87E+08	4398000	7.92E+08	126094613.4
x12	501	20507100	3212900	23720000	3258093.982
x13	501	27100000	2500000	29600000	4158071.6
x14	501	1.76E+08	5500000	1.81E+08	25350214.76
x15	501	5.26E+10	5500000	5.26E+10	7313927635
x16	501	89832300	5548700	95381000	20853799.92
x17	501	91317000	6003000	97320000	19977566.47
x18	501	50927600	1856400	52784000	8730773.196
x19	501	10565000	400000	10965000	1520614.987
x20	501	2.48E+08	943160	2.49E+08	44729707.35
x21	501	3.01E+08	11617000	3.13E+08	39854985.18
x22	501	66577200	1102800	67680000	15197800.36
x23	501	5.54E+09	3736600	5.54E+09	807533642.1
x24	501	48063980	2023900	50087880	8734727.73
x25	501	1.74E+08	6522200	1.81E+08	25590983.93
Log output	501	4.1376	7.7724	11.91	0.48331841

Table 2. Regression coefficients values

Model	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
Constant	9.194	0.11		83.906	0
x1	1.15E-08	0	0.218	3.188	0.002
x3	-3.86E-08	0	-0.745	-14.132	0
x4	4.31E-09	0	0.513	5.824	0
x6	-4.58E-08	0	-2.094	-13.778	0
x7	1.23E-08	0	0.908	7.616	0
x8	1.08E-09	0	0.257	4.666	0
x9	8.25E-09	0	0.659	9.795	0
x10	4.02E-09	0	1.17	6.747	0
x11	-1.18E-09	0	-0.836	-7.103	0
x12	-4.64E-08	0	-0.647	-7.546	0
x13	-9.49E-08	0	-2.119	-23.481	0
x14	-3.43E-09	0	-0.2	-5.163	0
x15	-6.62E-11	0	-0.579	-26.651	0
x16	-8.45E-09	0	-0.753	-5.869	0
x17	1.44E-08	0	0.451	20.213	0
x18	-4.43E-09	0	-0.051	-4.849	0
x19	2.17E-07	0	0.325	18.262	0
x20	1.53E-09	0	0.184	9.124	0
x21	-1.11E-08	0	-0.331	-12.207	0
x22	2.32E-08	0	0.611	19.678	0
x23	1.36E-10	0	0.423	15.483	0
x24	2.69E-08	0	0.284	18.415	0
x25	2.57E-08	0	4.405	18.961	0

dent variables was effective according to statistical analysis and the P-values where P was less than or equal to 5% for all variables. The values of un standardized coefficients in Table (2) were used to build a prediction model:

$$y=[9.194 + 1.146*10^{-8}x1-3.860*10^{-8}x3 + 4.310*10^{-9}x4-4.579*10^{-8}x6+1.231*10^{-8}x7+1.080*10^{-9}x8 +8.252*10^{-9}x9+4.017*10^{-9}x10-1.177*10^{-9}x11-4.635*10^{-8}x12-9.49*10^{-8}x13-3.434*10^{-9}x14-6.618*10^{-11}x15-8.45*10^{-9}x16 + 1.435*10^{-8}x17-4.432*10^{-9}x18 + 2.16*10^{-7}x19+1.529*10^{-9}x20-1.109*10^{-8}x21+2.324*10^{-8}x22 + 1.361*10^{-10}x23+ 2.687*10^{-8}x24\ 2.573*10^{-8}x25] \quad (2)$$

This model removed the landfill item (x2) and the construction work above moisture proof layer (x5) parameters because of their unimportance. Table 3 displays the results of the statistical measures that use to prove the validity of the regression techniques. Where the value of (R) equal to 99.8%. The positive value refers to the strong correlation between a dependent variable (actual cost) and an independent variable.

Also, the (R^2) equal to 99.6% which means that 99.6% of the total variation in dependent variable (actual cost) (y) can be explained by the linear relationship between *xi* and *y* (as described by the regression equation).

Table 3. Results of the correlation

Model				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.998	.996	.996	1.124051685

7. Model Validation

Model validation is very important step in building a cost model to test its accuracy it includes testing and evaluating the developed model with some validation or test data. The validations data is taken randomly from the data set and should not enter in model develop. The researcher used 30% of data to cheek accuracy of model. To evaluate the validity of the derived equation of the model for the final cost of construction projects, the natural logarithm (Ln) of prediction cost is draw beside the natural algorithm (Ln) of real cost for test data set as shown in Figure 1.

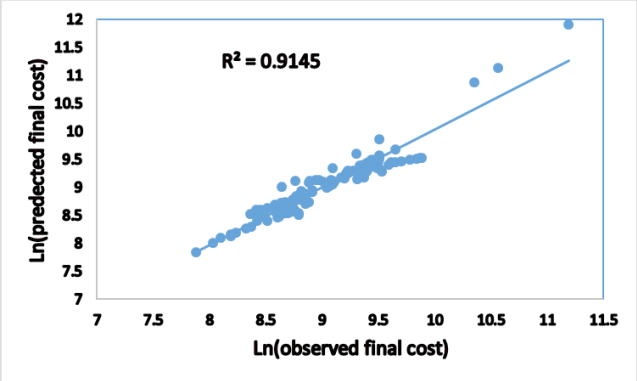

Figure 1. Predicted value versus observed final cost.

The coefficient of determination (R^2) is obtained to be (91.8%), so it can be decided that this model shows a very good agreement with the real observations.

The same statistics parameters that use[13] is used in this research to establish the average accuracy of developed model, as shown in the Table 4, where the Mean Absolute Percentage Error (MAPE) equal to 1.03% that mean the Average Accuracy (AA) of this model equal to 98.8 % is very good as shown in the Table 4.

Table 4. Statistical measures result for regression model

Description	Statistical Parameters
MAPE	1.03%
AA	98.97%
R	95.80%
R^2	91.80%

8. Conclusion

The researcher concludes some point from study as shown below:

1. The regression analysis technique proved its ability for prediction of the budget of construction Projects. One prediction model is built and the result showed that it is very accurate as the degree of accuracy was 98.97%.
2. Coefficient of determination is utilizing for determining the linear relationship between real cost and predict cost with 91. 80%.
3. The developed model can be used by the stakeholders to predict the cost of construction since it is simple and easy to use.

9. Proposals for Future Studies

1. Replace regression analysis technique with other techniques such as support vector machine technique and make comparison between the two techniques to show the best technique in terms of degree of accuracy for using in construction management field.
2. Use additional items of construction project to improve the prediction model across a larger range of data.
3. Use multiple linear regressions to develop models for items of project use to predict the cost index of items.

10. References

1. Amin A. Time-cost-quality-risk of construction and development projects or investment. Middle-East Journal of Scientific Research. 2011; 10(2):218–23.
2. Cheng MY, Hsing CT, Erick S. Conceptual cost estimates using evolutionary fuzzy hybrid neural network for projects in construction industry. Expert Systems with Application. 2010; 37(6):4224–31. https://doi.org/10.1016/j.eswa.2009.11.080
3. Al-shanti YA. A cost estimate system for gaza strip construction contractors, Palestine. PhD Thesis. Master Thesis in Construction Management, The Islamic University of Gaza Strip; 2003. p. 1–191.
4. Ayed AS. Parametric cost estimating of highway projects using neural networks. Master, Faculty of Engineering and Applied Sciences, Memorial University, Newfoundland; 1997. p. 1–102. PMid:9242309
5. Al-Zwainy FMS, Hadhal NT. Investigation and evaluation of the cost estimation methods of Iraqi communication projects. International Journal of Applied Engineering Research Management. 2015; 5(6):41–8.
6. Preliminary cost estimating models for road construction activities [Internet]. [cited 2010 Apr 16]. Available from: https://www.researchgate.net/

publication/238746511_Preliminary_Cost_Estimating_Models_for_Road_Construction_Activities.

7. Lowe DJ, Emsley MW, Harding A. Predicting construction cost using multiple regression techniques. Journal of Construction Engineering and Management. 2006; 132(7):750–8. https://doi.org/10.1061/(ASCE)0733-9364(2006)132:7(750)
8. Al-Zwainy FMS, Abdulmajeed MH, Aljumaily HSM. Using multivariable linear regression technique for modeling productivity construction in Iraq. Open Journal of Civil Engineering. 2013; 3(3):127–35. https://doi.org/10.4236/ojce.2013.33015
9. Al-Zwainy FMS, Hadhal NT. Building a mathematical model for predicting the cost of the communication towers projects using multifactor linear regression technique. International Journal of Construction Engineering and Management. 2016; 5(1):25–9.
10. How to Choose the Right Forecasting Technique [Internet]. [cited 1971]. Available from: https://hbr.org/1971/07/how-to-choose-the-right-forecasting-technique.
11. NIST/SEMATECH e-Handbook of Statistical Methods [Internet]. [cited 2012]. Available from: https://www.itl.nist.gov/div898/handbook/index.htm.
12. Using multivariate statistics [Internet]. [cited 2018 Jul 02]. Available from: http://www.mypearsonstore.com/bookstore/using-multivariate-statistics-0134790545.
13. Khaleel TAM. Development of the artificial neural network model for prediction of Iraqi expressways construction cost. International Journal of Civil Engineering and Technology. 2015; 6(10):62–76.