Building a Mathematical Model for Predicting the Cost of the Communication Towers Projects Using Multifactor Linear Regression Technique

Faiq Mohammed Sarhan Al-Zwainy[1,*], Neran Taher Hadhal[2]

[1]Department of Civil Engineering, Alnahrain University, Baghdad, Iraq
[2]Department of Civil Engineering, Baghdad University, Address, Baghdad, Iraq

Abstract   Due to lack of data and information in feasible study of the construction projects, therefore the predicating of the construction cost is very difficult. The objective if in this study is development a mathematical model for predicting the cost of the communication towers projects. Multifactor linear regression technique is developed and used for predication of the cost of communication towers projects in Iraq. Seven effectiveness factors are used for cost forecasting by MLR model, they involve Security Conditions, Tower Types, Experience of Contractor, Foundation Types, Tower High, Main Cable and Site Area. It was found that Multifactor linear regression has the flexibility to foretell the cost with a excellent degree of accuracy 90.1%, mean absolute percentage error 9.891% and coefficient of correlation (R) was 98.6%.

Keywords   Communication Tower Projects, Cost Estimation, MLR Model, SPSS

1. Introduction

Cost prediction remains a difficult and complex problem, in spite that, the researchers were still studying and trying different approaches and methodologies to solve it. Preparation of a construction cost estimate for any project is a very complex process. Process of construction cost estimation contains many variable factors. Every variable has to be correctly estimated based on proper study, past experience and research to calculate total project cost of construction [1].

The researcher believed the Prediction of the cost estimation of communication towers projects is a vital mission in the management of Iraqi construction sector. The scope of this research is to build and evaluate an analytical model which may be used to estimate the budget of towers projects, by utilizing Multifactor Linear Regression Technique (MLRT), through the following steps:

- Identification of MLR model variables, which have an effect on the cost of communication towers projects in Iraq.
- Build an inclusive tool for parametric cost estimation utilizing the MLR as a favorable predicting model.
- Develop a numerical equation to calculate the cost of communication towers projects.

- Investigation the verification and validation of the mathematical model developed.
- Determine the degree of accuracy for the numerical equation and clarifying the amount of the correlation between predication costs and actual cost.
- Investigation and assess of the proposed Multifactor Linear Regression models.

2. Research Methodology

Research methodology has adopted to involve the following:

- Literature Survey: is to review the cost estimation and application multifactor linear regression technique in communication project. The concept of the cost estimation and regression analysis, benefits, structures, and applications of regression analysis in project management field, Which include the review of literatures involving references, thesis, papers, books and web-site relating to the subject of research especially which are related to construction industry.
- Field Work: This involved data description and selection of the variables, then development of Multifactor Linear Regression (MLR) Model. And validity of the MLR Model.

3. Selection Variables

In this study, the researcher used quantitative and

qualitative approaches to collection data, as shown following:

1) The qualitative approach: in this approach, interviews and the questionnaire were used to investigation and determine the main factors affecting the cost of communication towers projects in Iraq. Most effective factors can be shown in Table 1. And, the detail of this approach was explained in manuscript (Al-Zwainy and Hadhal, 2015) [1].

2) The quantitative approach: was used to gather the real data from resources (Asiacell and Zain companies) by filling a form for each communication towers project, which contains the input factors and the actual cost for each project as output. Multifactor Linear Regression model require a lot of historical data. All of this communication towers projects were done between 2013 and 2014 in Iraq. The projects information was gathered from direct watching and reports of contractors and companies. This method faces a great difficulty nowadays because of the unsecured status of the country (Iraq), and the shortage in documentation. In spite of these obstacles, the researcher succeeded in gathering well trusted data for more than 45 communication projects through the companies' visits and reading the concerned sheets, contract documents and reports of projects.

Table 1. Effective Factors the Cost of Communication Towers Projects in Iraq

Variable	RM	Important
X_1	Site Works or Site Area	Important
X_2	Foundation Types	Important
X_3	Tower Types	Important
X_4	Tower High	Important
X_5	Main Cable	Important
X_6	Section Type	Important
X_7	Method of Assembling the Tower	Non important
X_8	Site Conditions	Non important
X_9	Quality of Completed Works	Important
X_{10}	Market Condition	Important
X_{11}	Quality of Planning and Control	Non important
X_{12}	Availability of Materials	Non important
X_{13}	Security Conditions	Important
X_{14}	Holidays and Emergency events	Non important
X_{15}	Cost of Construction Materials	Non important
X_{16}	Variation Orders	Non important
X_{17}	Experience of Contractor	Important
X_{18}	Classification of company	Non important
X_{19}	Construction Method	Important
X_{20}	Efficient of cost management	Non important

These variables are represented fell into two classifications, subjective and objective variables. All variable can be analyzed so as to find the optimum approach to illustration for the variable within the modeling process, as shown in Table 2.

Table 2. Classification of variables

Variable	Classification of variables	Description	Units
V1 (x13)	subjective variables	Security Conditions	1=Stable 2 = Unstable
V3 (x3)		Tower Types	1 = Roof Top 2 = Green Field
V6 (x17)		Experience of Contractor	1 = Good 2 = Bad
V2 (x2)	objective variables	Foundation Types	Cubic meter
V4 (x4)		Tower High	meter
V5 (x5)		Main Cable	meter
V7 (x1)		Site Works or (Site Area)	Square meter

Objective variables content four variables, and may be measured utilizing the unit of measurement such as High of Tower (V4) and Length of Main Cable (V5) were measured in length meter. While the second variable is Foundation Type (V2) measured in cubic meter, and the variable offsite Area (V7) can be measured by square meter.

Subjective variables content three variables, and could be measured depend on the coding system, such as, security conditions (V1) will be classifies to security conditions and non-security conditions and specify them the value 1 and 2, respectively. Also the third variable Tower Type (V3) specifies the nature of tower as green filed or roof tower. It assigns them the values of 1 and 2 respectively. Latterly, the sixth variable Experience of Contractor (V6) defines an index which accounts for Experience of Contractor with two types good, and poor. It specifies them the values of 1 and 2 respectively.

4. Development of Multifactor Linear Regression (MLR) Model

A lot of many subscriptions and books described the multiple linear regressions technique. The researcher founded the text by (Al-Zwainy et al., 2013) [2] and (Smith, 1999) [3] could be particularly useful because it is simply and understand. In this study, multifactor linear regression can be used as an optimum forecasting model and very strong mathematical tool that allows the engineers and researchers to find the engineering relationship between dependent variables (e.g. actual cost) and independent variables (objective and subjective variables) (e.g. Foundation type, tower type, etc.) It can tentatively formulate (Johnson and Bhattacharyya, 2006) [4]:

$$Y_i = \beta_0 + \beta_1 \chi_{i1} + \beta_2 \chi_{i2} + \ldots\ldots + e \qquad (1)$$

Where:

● $i = 1,\ldots,n$
● x_{i1} and x_{i2} are the values of the input variables for the i the experimental run
● Yi is the corresponding response
● e_i the error components are assumed to be independent normal variables with mean zero and variance σ^2
● $\beta_0, \beta_1, \ldots, \beta_p$ the regression parameters are unknown and so is σ^2

In the field of scientific research there is much statistical software. But in Iraq, the most common software is Statistical Package for the Social Sciences (SPSS). The developers of the (SPSS) made every effort to make the software easy to use. Also, the researcher believed that SPSS is one among the analytics software's are estimate with confidently what will happen next and will be create smarter choices, improve outcomes and solve complex problems. In this study, the researcher used SPSS Version. 24 to development of the multifactor linear regression model and statistical analysis results as shown in Table 3., Table 4., and Table 4.

Questionnaire forms were completion and distribution, then analyzing those statistically according statistical lows. Summary of statistical analysis can be seen in Table 3.

The quantity (R), called the linear coefficient of correlation, measures the direction and the strength of a linear relationship between two variables. And the determination coefficient (R^2) represents the percent of the data that is the nearest to the line of best fit. Table 4 show results of the correlation test, where the value of coefficient of correlation equal to 99.2%. The positive value means the correlation between a dependent variable (actual cost) and an independent variable is very strong. Also, the Coefficient of determination equal to 98.4% means that 98.4% of the total variation in dependent variable (actual cost) (y) will be explained with the linear relationship between x_i and y. The other 1.6% of total variation in y remains unexplained.

Table 5. shows the regression variables, regression coefficients and statistical tests of the P-values of coefficients.

Table 3. Descriptive Statistics

variables	N	Range	Minimum	Maximum	Std. Deviation
Security conditions	45	1	1	2	0.4995
Foundation type	45	157	3	160	34.991
Tower Type	45	1	1	2	0.4903
tower High	45	58	12	70	15.7545
Maine cable (Length)	45	118	12	130	33.2695
Experience of Contractor	45	1	1	2	0.4841
Site Works (Site Area)	45	430	70	500	153.0535
Actual Cost (ID)	45	157	18.5	176	45.1498

Table 4. Results of the Correlation

Model	Model Fit statistics			
	R	R-squared	RMSE	MAPE
actual cost -Model_1	0.992	0.984	5.755	10.277

Table 5. Regression Coefficients Values

Coefficients								
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
	B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	32.712-	6.154		-5.315-	.000	-45.182-	-20.242-
	Security conditions	3.018	2.221	.033	1.358	.183	-1.483-	7.518
	Foundation type	.377	.065	.292	5.795	.000	.245	.509
	Tower Type	18.420	3.902	.200	4.721	.000	10.515	26.326
	High tower	.922	.189	.322	4.873	.000	.539	1.306
	Main cable	.091	.033	.067	2.737	.009	.024	.158
	Experience of contractor	4.579	2.612	.049	1.753	.088	-.713-	9.870
	Site work	.065	.019	.221	3.439	.001	.027	.103

a. Dependent Variable: total cost

Regression analysis generates an equation to explain the mathematic relationship between one or a lot of predictor variables and also the response variable. SPSS software labels the standardized regression coefficients as "Beta" and unstandardized coefficients are labeled "B". Coefficient Standardization is sometimes done to answer the question of that of the independent variables have a greater impact on the dependent variable in a multiple correlation analysis, once the variables are measured in numerous units of measurement. Key of the coefficients understanding is slopes, and they are sometimes called slope coefficients. Can be concluded from this table, independent variables (foundation type, tower type, height of the tower, the main cable and site location) were significant from a statistical standpoint, according to T-test (at the P-values P≤ 0.05). But the independent variables (security conditions and the Experience of the contractor) did not have a significant impact on the multiple regression models, according to T-test.

The researcher believed a regression models as a powerful tool will be utilized to estimate the cost of communication tower project from multi variables, and it allowing to prediction the cost of projects in future depending on the information about past or present projects. Values of unstandardized coefficients in Table 5 can be used to build the multivariable linear regression, as shown the following eqution.

$$TotalCost = -32.712 + 3.018SC + 0.377FT$$
$$+ 18.42TT + 0.922HT + 0.091LC \quad (2)$$
$$+ 4.579EC + 0.065SW$$

5. Validity of the Multiple Linear Regression MLR Model

The purpose of the validation is to provide a statistically reliable model. In this study, twelve new documents for communication tower projects were gathering as shown in Table 6. These documents weren't included in model developed, but utilized for validation and verification of this model. Actual cost of communication tower projects and the predicted values (by suggested estimation equation) were presented in Table 7.

In order to prove the validity of the model developed, the researcher used the coefficient of determination to explain the linear relationship between projected cost and actual cost, where the coefficient of determination equal to 97.32%, as shown in Figure 1 which means that model showing the excellent homogeneous with actual observations.
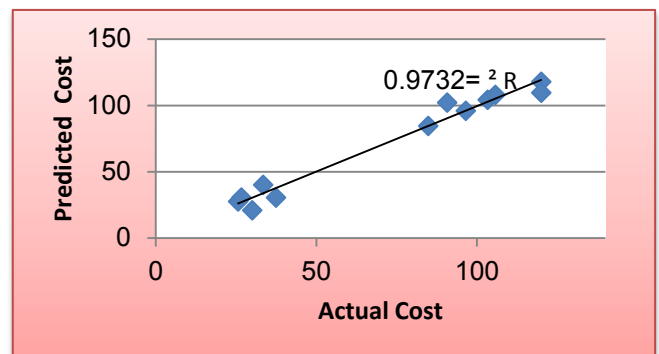
The researcher used the same statistics measures adopted by (Al-Suhaily et al., 2015, Al-Zwainy, 2014, and Al-Zwainy et al., 2012) [5-7] to determine the average accuracy of the this model, as shown in the Table 8 below, where the Mean Absolute Percentage Error (MAPE) equal to 9.891% that mean the Average Accuracy (AA) of this model equal to 90.1% is excellent.

Table 6. Information of the New Communication Tower Projects

Project No.	Security conditions	Foundation type	Tower Type	Tower High	Maine cable (Length)	Experience of Contractor	Site Works or (Site Area)	Actual Cost (ID* 10^6)
P1	1	50	2	35	120	2	400	103.3
P2	1	80	2	45	130	1	350	120
P3	2	3	1	13	100	2	70	25.6
P4	1	6	1	22	100	2	150	33.4
P5	2	48	2	40	75	1	300	96.5
P6	2	26	2	50	75	2	400	105.7
P7	1	5	1	27	65.3	1	70	37.4
P8	2	63	2	30	60	1	200	84.8
P9	1	60	2	40	55	1	400	90.7
P10	2	9	1	18	50	2	80	26.6
P11	1	4	1	14	40	2	80	30
P12	1	75	2	50	13	1	350	120

Table 7. Actual Cost and the Predicted Cost

Project No.	Actual Cost	Estimate Cost	MPE
P1	103.3	104.35	-1.01646
P2	120	117.96	1.7
P3	25.6	27.67	-8.08594
P4	33.4	40.2	-20.3593
P5	96.5	96.05	0.466321
P6	105.7	108.05	-2.22327
P7	37.4	30.58	18.23529
P8	84.8	84.62	0.212264
P9	90.7	102.23	-12.7122
P10	26.6	30.64	-15.188
P11	30	21	30
P12	120	109.66	8.616667
Σ			-0.35461

Figure 1. Validation MLR Model

Table 8.   Statistical Measures Results for Regression model

Description	Statistical parameters
RMSE	4.72%
MAPE	9.891%
AA%	90.1%
R	98.6%
R^2	97.32%

6. Conclusions

Traditional of estimation methods suffer from the major disadvantages of the lack of precision and uncertainty, but a regression analysis offer several advantages over traditional methods for the prediction of the communication projects cost. The conclusions drawn from the above study were summarized below:

- Multifactor linear regression has the ability to estimate the cost of communication towers Projects, one an optimum forecasting model Developed as a comprehensive tool for parametric cost estimation with high degree of accuracy with 90.1%.
- Coefficient of determination using to explain the linear relationship between projected cost and actual cost with 97.32%
- The developed model is useful for researchers and stakeholders because it provides easy tool to predict the cost of construction Projects.
- The developed model on preparing the historical data equal to 45 past project as historical data to be modelled and determining the relation between output and input variables. Seven influencing factors are used for cost forecasting by Multifactor linear regression model, they include Security Conditions, Tower Types, Experience of Contractor, Foundation Types, Tower High, Main Cable and Site Area.

7. Proposals for Future Studies

The researcher here introduces a number of proposals to be investigated as future studies and as follows:

- Repeating the same study by using neural network technique and making a comparison between the results

to show which method is more adequate for use in construction management.
- Developing several MLR models to demonstrate the ways in which different types of civil engineering problems ensure the successful development and application of this technology to project management problems.

REFERENCES

[1] Al-Zwainy, F. M. S. and Hadhal, N. T. (2015). "Investigation and Evaluation of the Cost Estimation Methods of Iraqi Communication Projects" International Journal of Engineering and Management Research, IJEMR, Vol. 5, No. 6, pp. 41-48. http://www.ijemr.net/DOC/InvestigationAndEvaluationOfTheCostEstimationMethodsOfIraqiCommunicationProjects(41-48).pdf.

[2] Al-Zwainy, F. M. S., Abdulmajeed, M. H. and Aljumaily, H. (2013). "Using Multivariable Linear Regression Technique for Modeling Productivity Construction in Iraq" Open Journal of Civil Engineering, OJCE, Vol. 3, No. 3, pp. 127-135. doi:10.4236/ojce.2013.33015.

[3] Smith, S. D., (1999) "Earthmoving Productivity Using Linear Regression Techniques," Journal of Construction Engineering and. Management, Vol. 125, No. 3, pp. 133-141. doi:10.1061/(ASCE)0733-9364(1999)125:3(133).

[4] Johnson, R. A., Bhattacharyya G. K., (2006). Statistics: Principiles and Methods, 5ed, John Wiley and Sons, Inc., USA.

[5] Al-Suhaily, R. H., Saco, Z. M., Alzwainy, F. M. S., (2015). Project Management and Artificial Neural Networks: Fundamental and Application, LAP LAMBERT Academic Publishing, Germany.

[6] Al-Zwainy, F. M. S. (2014). "Development of the Mathematical Model for Predicating the Construction Productivity in Iraq Using the Artificial Neural Perceptron Network" Journal of Engineering and Development, Vol. 18, No. 2, pp. 1-21.

[7] Al-Zwainy, F. M. S., Hatem, A. R. and Huda, F. I., (2012). "Using Artificial Neural Network for Finishing Works for Floors with Marble," ARPN Journal of Engineering and Applied Sciences, Asian Research Publishing Network (ARPN) Vol. 7, No. 6, 2012, pp. 714-722. http://www.arpnjournals.com/jeas/research_papers/rp_2012/jeas_0612_714.pdf.