TJES
Tikrit Journal of Engineering Sciences

Salar S. Ahmed[*1]
Noori S. Ali[2]

[1]Civil Engineering Department, Engineering College, Salahaddin University, Erbil.
[2]Civil Engineering Department, Engineering College, Cihan University, Erbil.

Pre-Design Cost Modeling of Road Projects

A B S T R A C T

The conceptual cost estimation is a critical indicator of the future of the project's success at the initial process before submitting the project design. Pre-Design cost measuring is recognized as the very important procedure which affects the future of constructing proposed projects of roads. Pre-design cost is a preliminary cost measuring and is considered the prediction of the cost of a project during the planning and design phase. This evaluating and measure provides a great foundation for the process of financial support decisions and cost control. The aim of this study is to create pre-design cost measuring model for constructing road projects by utilizing linear regression technique. The research methodology consists of the data collected from public sectors which are the actual cost of the constructed road projects in Erbil governorate, then models have been developed by applying SPSS software, then the models summarized and selected. The mean absolute percentage error (MAPE) has been calculated using Excel program to calculate degree of the accuracy for the produced models. The accuracy was between -29% and +51%. As a result of the study regression models are useful and has a big advantage in predicting project cost in pre-design process and the planning stage of the project, by using either simplex computer scheme or any measuring tools.

نماذج تخمين التكلفة لمرحلة ما قبل التصميم لمشاريع الطرق

سالار صالح أحمد / قسم الهندسة المدنية، كلية هندسة ، جامعة صلاح الدين ، اربيل ، العراق
نوري صادق علي/ قسم الهندسة المدنية، كلية هندسة ، جامعة صلاح الدين ، اربيل ، العراق

الخلاصة

يعد مفهوم تقدير التكلفة مؤشرا حاسما لمستقبل نجاح المشروع في المراحل الأولية قبل تقديم تصميم المشروع. يعد تقدير التكلفة في مرحلة ما قبل التصميم أهم عملية تؤثر على مستقبل مشاريع بناء الطرق الجديدة. إن تقدير التكلفة قبل التصميم هو حساب التكلفة ألاولية للمشروع ويعد بمثابة التنبؤ بتكلفة المشروع أثناء مرحلة التخطيط والتصميم. يوفر هذا التقييم والحساب أساسًا عظيما لعمليات اتخاذ القرارات المتعلقة بالدعم المالي وضبط التكاليف. الهدف من هذا البحث هو إنشاء نماذج لتقدير تكلفة المشاريع في مرحلة ما قبل التصميم لمشاريع إنشاء الطرق من خلال استخدام تقنية الانحدار الخطي. تتكون منهجية البحث من جمع البيانات للتكلفة الحقيقية من القطاعات العامة لمشاريع الطرق الخاصة بمدينة أربيل ثم بعد ذلك تم تطوير النماذج من خلال تطبيق برنامج SPSS ، ثم تلخيص النماذج واختيار الأفضل. تم حساب متوسط خطأ النسبة المئوية المطلقة (MAPE) باستخدام برنامج Excel لقياس دقة النماذج المطورة. كانت الدقة بين 29%- و51%+ تبين من خلال النتائج الحاصلة من الدراسة أن نماذج الانحدار الخطي مفيدة وتتجلى فائدتها في تقدير التكلفة الأولية للمشاريع لمراحل التخطيط وما قبل التصميم إما باستخدام برنامج كمبيوتر بسيط أو آلة حاسبة.

الكلمات الدالة: الدقة، الكلفة، تخمين، نموذج، تصميم مسبق، انحدار.

*Corresponding Author: salarsalih1975@gmail.com

1. Introduction

The industrial sector is greatly affected by poor business cost planning in relations of budget overruns, which have significant negative outcomes for projects [1]. Pre-design cost is a preliminary cost measuring and is considered the prediction of the cost of a project during the planning and design phase. This evaluating

and measure provides a great foundation for the process of financial support decisions and cost control [2]. Pre-Design cost measuring is recognized as the very important procedure which affects the future of the construction of new road projects. Estimating is considered as the kernel of the engineering construction work cost and as a result proper interests will be received over the years [3].

The conceptual cost estimation is a critical indicator of the future of the project success at the initial process before submitting the project design. It is a process based on the background, which enters the extraction of different relationships between cost and its effect factors [4].

Completion project in accordance to the planned and scheduled time, cost and quality is the main goal of the pre design cost prediction, which is considered the important task of the estimator, to explore the best model with less error for the proposed project [5].

The method or the technique is utilized to measure the most important factor of the engineering construction management system which is the cost, from other physical factors of the system. The importance of the pre-design cost estimation is bringing out in giving owners the capability to take the correct decision concerning the feasibility of the proposed project by utilizing the accurate pre-design cost estimation process. The availability of the necessary information at the specified time is greatly determining the range of the pre-design estimation cost accuracy for the proposed project. Therefore, pre-design estimation cost availability is very important at the early stage where little information is available related to the proposed project. However, pre-design estimates are greatly required to construction engineering owners in order to check them prior to taking action to go on with a project [6].

During the conceptual process, whereas applying cost forecasting, many problems related to cost estimation will appear which are lack of pre-design data, shortage of information of road work costs and shortage of cost prediction techniques. More problems come because of the wide suspicions resulting from engineering resolutions, socio-economical, and environmental problems. Parametric cost predicting depending on historical data at the time of conceptual forecasting process which is always utilized in civilized countries. Accordingly, civilized countries confront problems concerning the construction of the road costs data, probably used for cost forecasting in both conceptual process and feasibility study of the project life cycle [7].

The term cost measuring of the project fund is considered the main significant issue for each of the project client and project executer who is usually a contractor [8].

Hackney [9] reported a detailed list of items with definitions and evaluation for the capital projects. The research supposed to utilize the defined item list for applying emergency to capital cost measure and then check accuracy through comparing the evaluated item list of thirty construction work to their specific grades of expenditure overrun.

Kim, Seo [10] utilized hybrid models of genetic algorithms (GA) and neural networks (NN) to anticipate cost of residential projects to forecast initial cost measure depending on past information of constructed projects for the period (1997-2000) in Seoul, Korea.

Lawther and Edwards [11] in their research indicated to regression analysis as a statistical technique in which historical data is evaluated in a structured format. The main parameters of construction are specified and a mathematical relationship between variables, in the format of an algebraic equation, is defined by utilizing historical data. From the 1970s till now continuous regression model is utilized for measuring cost. The aim of linear regression is to utilize the relation between dependent parameters, which is linear, and independent parameters to measure or define the behavior of the dependent parameters. Multiple regression equation usually appears in the form: $Y = C + A1X1 + A2X2 + \ldots AnXn$. Where Y is the total predicted cost; C is a constant or line intercept; and X1, X2, etc. are the value of parameters that might assist to measure y; and A1, A2 …etc. are the coefficients measured by regression operations. The regression analysis can then be utilized to estimate the magnitude of dependent parameters once the magnitudes of the independent parameters are entered. Merrow and Yarossi (1990) in their research used y as the measured cost/real cost and x parameter like degree of domain definition, and degree of invalidate technology [12].

Lowe, Emsley [13] in the United Kingdom used 286 sets of project data to improve linear regression models to estimate the engineering cost of the building projects. They specified forty one prospects independent parameters, and by applying regression analysis, illustrate 5 important affected parameters like total internal area of the floor, project assignment, time, mechanical setup, and pile placing.

Abu Jamous [14] in his research shows that traditional parametric cost measuring methods have been used by numerous researchers because of the easy format. Also in the U.K. residual building sector, an improvised iterative technique has been used to carry out the multi-variety regression analysis by improving two discrete models and suggested them for measuring the cost and time of the project. The percentage of the accuracy that has been carried out on the statistical operations reached to 81.4% and 92.7% for cost and time models respectively, the accuracy processes conducted with respect to adjusted coefficient of determination.

Skitmore and Ng [15] as well, performed the same way; by utilizing standard regression and a forward cross validation regression analysis for the improving of numerous models for estimating real construction duration and cost by utilizing a set of 93 Australian construction projects.

Mahamid, Brul [3] improved a linear regression model to estimate the cost of highway construction items depending on one hundred packages of information in the west Bank of Palestine. The estimation models were improved for 3 important highway construction items, which are excavation and

filling, base course works, and asphalt operations. 3 sets of models of every item were improved depending on submitted dependent parameters; they are whole expenditure of work item, cost for each meter length, and cost for each square meter. The suggested independent parameters are length of road, width of pavement, width of base course, topography, drill ability and suitability of the soil. The determination coefficients r^2 of the improved models where between 0.57 and 0.96.

Hegazy and Ayed [16] developed 12 normally, cost-estimating relationships using regression analysis on historical project data. The improvement of these models, nevertheless, is considered a complicated operation because of the natural restrictions of regression analysis.

Sodikov [7] stated that the World Bank transport section has conducted one of the earliest trials to improve estimation data information submitted in term of Road Costs Knowledge System (ROCKS) in order to use in the developed countries. Road Costs Knowledge System (ROCKS) attempts to found a model in order to improve a global information system on highway project cost so as to put a foundation of an organizational data base, to get the average and the limit of the unit price depending on past information that can increase and develop the accuracy of new cost forecast and decrease the risk made by cost overruns. Establishing this type of road data information organization in the developed countries helps to create resources and assist right assessment for the future road projects. Artificial neural networks (ANN) can be applied in order to use database with more sufficiency and to get solution for the matters ongoing studies to reveal the accessibility of cost forecast in the conceptual process.

Many researchers have defined the problems related to the cost estimation in the initial steps of project improvement. Hegazy and Ayed [16] utilized a neural network method to arrange construction project cost information and improve a parametric cost predicting model for road projects. They defined two substitutional methods to examine weight of networks: GAs (genetic algorithms) and simplex optimization using software (Excel's inherent solver function).

Adeli and Wu [17] showed that road construction costs vary and the variety is made from several unexpected variables like variables concern to persons' ideas and opinions, unexpected fluctuations in the market, and environment situation.

Kim, An [18] investigated various techniques of cost prediction models in the preliminary process of building work projects like multiple regression equation, case based reasoning and neural networks. They proved that neural networks generate better estimation accuracy while case-based reasoning showed best results for the long run.

Researchers improved a method for creating range measures to assess the risk of cost increasing in residual building projects by utilizing artificial neural networks. They summarized a process for building a prototype model for cost prediction, and using

Artificial neural networks model to create prediction for construction engineering projects [7].

Wilmot and Cheng [19] improved application of the model in estimating of road construction costs in Louisiana. They indicated that the model reflects approximately past construction costs from the year 1984 to1997.

They stated that the model expects future estimation of the road construction costs in Louisiana become twice for the period between 1998 and 2015. Submitting estimation to cost-cutting policies and supposing input costs are 20% less than expected. The models predict that road construction costs will rise by 75% for the period between 1998 and 2015.

Oberlender and Trost [20] emphasized more on developing forecasting accuracy by entering 4 determinants affecting the accuracy of initial cost predictions. Going on, they suggested a multiple regression equation which produces an important relation between forecast accuracy and effect variables like principle of design, team skills and cost data, time limitation to prepare the forecast, field requirements, and tendering process and labor situation.

The Korean Ministry of Construction and Transportation studied collaboration, examined the real budgeting plan in the road construction project. They improved the two-tiered cost forecast models of road construction projects, taking into consideration the aim of estimation, accuracy limit, and level of the data available in each step of project funding and initiation [6].

Due to lack of data and information and urgent needs for the cost at the initial stage of the proposed project; cost predicting model is highly required in order to estimate the pre-design cost for the phase of allocating funds to any road project of the public sectors. Therefore, the aim of this study is to create pre-design cost estimation model for the road construction projects (Two-Lane Two-Way Road) for the public sector by utilizing linear regression technique.

2. Research Methodology

After defining the aim of this study and considering the literature review, the research methodology includes the data collection and tabulation and organization. It involves also developing regression method, then summarizing models in order to select the suitable models.

2.1 Data Collection

In this study road projects data were gathered in executive public sectors. The data were collected from executive Ministry of Reconstruction and Works, Ministry of Municipality Affairs, and Ministry of Interior Affairs. The main data involved cost of project and project characteristic of length, width and quantities of earth and asphalt of 42 finished projects between years 2004 and 2015.

2.2 Developing Models

SPSS software version 24 has been used to develop regression models, data and variables have been defined, and a series of mathematical models was produced using the multiple and linear regression analysis techniques in order to create the preliminary cost estimate equations. The standard model is shown as below form by [21].

$$Y = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 \ldots\ldots \text{Eq. (1)}$$

Where:

Y: is the dependent variable which refers to the Total Cost of the project

X1, X2, X3, and X4: are the dependent variables, which refer to the length and width of road (m), pavement width (m) and pavement thickness (m) respectively.

β_o, β_1, β_2, β_3, β_4: are regression estimated parameters

3. Results and Discussion

When a package of probable predictors was defined for the road projects, mathematical models were produced using linear regression technique. And the results of the SPSS criterion for all tests prove the adequacy of the tests which are shown in the Table 1 below: -

Table 1
SPSS Criterions Results

N (Number of Projects)	42
R^2	0.86
Adjusted R^2	0.85
Durbin Watson	2.7
ANOVA Test	0.000

Collinearity Statistics (VIF)	1

The following sections illustrate four groups of regression models that are produced to estimate the total cost of the road projects as below: -

3.1 Cost estimating model depends on road length

In the first model, the total cost of the road projects in IQD used as the dependent variable and length of the road in meters used as the independent variable; Cost= f(RL). The regression model was developed and the result is shown in Table 2 below: -

Table 2
Total Cost – Road Length Model

Model #	Independent Variables	Coefficients	P-value	R^2
1	(Constant)	-26850023.27	0.898	0.86
	Road Length (m)	348226.33	0.000	

The prediction equation of total cost produced by SPSS software is in the form below:-

Total Cost (ID) = (348226.33*Road Length) - 26850023.27……Eq. (2)

3.2 Cost estimating model depends on road length and width

In the second model, the total cost of the road projects in IQD used as the dependent variable and length of the road, width of the road in meter used as the independent variables; Cost= f (RL, RW). The regression model was developed and the result is shown in Table 3 below: -

Table 3
Total Cost – Road Length and Width Model

Model #	Independent Variables	Coefficients	P-value	Adjusted R^2
2	(Constant)	61815392.91	0.922	0.85
	Road Length (m)	347723.95	0.000	
	Road Width (m)	-7200253.85	0.881	

The prediction equation of total cost produced by SPSS software is in the form below: -

Total Cost (ID) = (347723.95*Road Length) - (7200253.85*Road Width) +61815392.91…Eq. (3)

3.3 Cost estimating model depends on road length, road width and pavement width

In the third model, the total cost of the road projects in IQD used as the dependent variable and road length, road width and pavement width in meter used as the independent variables; Cost= f(RL,RW,PW).

The prediction equation of total cost produced by SPSS software is in the form below: -

Total Cost (ID) = (353175.21*Road Length) - (122529209*Road Width) + (218566986.2*Pavement Width) - 308421401.6……Eq. (4)

The regression model was developed and the result is shown in Table 4 below:-

Table 4
Total Cost - Road (Length, Width and Pavement width) Model

Model #	Independent Variables	Coefficients	P-value	R² Adjusted
3	(Constant)	-308421401.600	0.644	0.85
	Road Length (m)	353175.210	0.000	
	Road Width (m)	-122529209.000	0.182	
	Pavement Width (m)	218566986.200	0.142	

3.4 Cost estimating model depends on road length, road width, pavement width and pavement thickness.

In the fourth model, the total cost of the road projects in IQD is used as the dependent variable and road length, road width, pavement width and pavement thickness in meter are used as the independent variables; Cost= f (RL, RW, PW, PT). The regression model was developed and the result is shown in Table 5 below: -

Table 5
Total Cost - Road (Length, Width, Pavement Width and Thickness) Model

Model #	Independent Variables	Coefficients	P-value	R² Adjusted
4	(Constant)	-367979560.200	0.687	0.85
	Road Length (m)	352433.30	0.000	
	Road Width (m)	-124878164.300	0.194	
	Pavement Width (m)	221111562.700	0.149	
	Pavement Thick. (m)	664737829.300	0.923	

The prediction equation of total cost produced by SPSS software is in the form below: -

Total Cost (ID) = (352433.3*Road Length) - (124878164.3*Road Width) +(221111562.7*Pavement Width) + (664737829.3* Pavement Thick.) - 367979560.2……Eq. (5)

3.5 Mean Absolute Percentage Error (MAPE)

The mean absolute percentage error (MAPE) has been calculated using Excel program to calculate the accuracy of the produced models. The below equation is used to calculate the MAPE [13].

$$\text{MAPE} = (1/N) \sum \left| (A_i - P_i)/ A_i \right| ……\text{Eq. (6)}$$

Where;

A_i: refers to the Actual Cost;
P_i: refers to the Predict Cost;
N: refers to number of projects.

Table 6 illustrates a summary of the tested regression model accuracy percentage (error %) that have been measured by Excel program.

Table 6
Summary of the Produced Regression Models with MAPE

Model #	Regression Model	MAPE %
1	Total Cost (IQD) = (348226.33* R. Length) - 26850023.27	±29.91
2	Total Cost (IQD) = (347723.95* R. Length) - (7200253.85*R. Width) +61815392.9	±32.54
3	Total Cost (IQD) = (353175.21*R. Length) - (122529209*R. Width) + (218566986.2*Pave. Width) - 308421401.6	±50.98
4	Total Cost (IQD) = (352433.3*R. Length) - (124878164.3*R. Width) + (221111562.7*Pave. Width) + (664737829.3* Pave. Thick.) - 367979560.2	±51.68

Best model is model #2 it can be applied to estimate cost for the new Two-Lane Two-Way Road projects.

Total Cost (IQD)=(347723.95* Road Length)-(7200253.85*Road Width)+61815392.9…Eq(3) .

4. CONCLUSIONS

According to the study, models can be produced in a regression form that could be utilized for pre-design cost measuring. Because of its easiness and capability to be treated by a simplex computer scheme or any measuring tools, these kinds of models are useful, and have a big advantage in predicting the cost of the projects in pre-design process and the planning stage of the project.

Because of the regression models are a prediction operation, therefore the accuracy of the results is not out of error and the errors were between -29% and +51%, consequently it is recommended to choose models #2 which depends on road length and width in case to measure total cost for projects of the roads.

The value of coefficients (R^2) and (adjusted R^2) for the produced models are between 0.85 and 0.86 which shows the strong relation between the dependent and independent parameters of the produced models and the expected magnitude of predicted models conform to the real-life data.

The results of the study conform with antedate studies and had recognized that the measured accuracy in the pre-design process of a project is ranging from ±25% to ±50%.

Length of the road (p-value=0.000<0.05) is affecting more to estimate the total cost of the project as SPSS's test results compared with other variables which have (p-value >0.05).

REFERENCES

[1] Saeed, Y.S., Cost and Time Risk Management in Construction Projects. *Tikrit Journal of Engineering Sciences*, 2018. 25(1): p. 42-48..

[2] Alumbugu, P.O., et al., Assessment of the factors affecting accuracy of pre-tender cost estimate in Kaduna state, Nigeria. *JOSR Journal of Environmental Science, Toxicology and Food Technology (IOSR-JESTFT)*, 2014. 8(5): p. 19-27

[3] Mahamid, I., A. Brul, and A. Bruland, Preliminary cost estimating models for road construction activities. *FIG Congress 2010 Facing the Challenges – Building the Capacity Sydney, Australia*, 11-16 April 2010, 2010.

[4] Adel, K., et al., Developing Parametric Model for Conceptual Cost Estimate of Highway Projects. *International Journal of Engineering Science*, 2016. 6(7): p. 1728-1734.

[5] Deshpande, P.V., Construction management: preliminary cost estimate and scheduling of MIT's *Civil and Environmental Engineering building*. 1999, *Massachusetts Institute of Technology*.

[6] Wali KI, Saber NI. Appraisal of Risk Factors which Influence the Construction of the School Buildings in Northern Iraq. *Tikrit Journal of Engineering Sciences* 2019; 26(1): 1-8.

[7] Sodikov, J., Cost estimation of highway projects in developing countries: artificial neural network approach. *Journal of the Eastern Asia Society for Transportation Studies*, 2005. 6: p. 1036-1047.

[8] Kim, D.Y., B. Kim, and S.H. Han. Two-staged early cost estimation for highway construction projects. in ISARK, *The 25th International Symposium on Automation and Robotics in Construction*. Institute of Internet and Intelligent Technologies. 2008.

[9] Hackney, J.W., Applied contingency analysis. *AACE Transactions*, 1985: p. 1-4.

[10] Kim, G., D. Seo, and K.I. Kang, Hybrid models of neural networks and genetic algorithms for predicting preliminary cost estimates. *Journal of Computing in Civil Engineering*, 2005. 19(2): p. 208-211.

[11] Lawther, P.M. and P.J. Edwards, Design cost modelling-the way forward. *Construction Economics and Building*, 2001. 1(1): p. 32-42.

[12] Baccarini, D. Estimating project cost contingency-Beyond the 10% syndrome. in Australian *Institute of Project Management National Conference*. 2005. Australian Institute of Project Management.

[13] Lowe, D.J., M.W. Emsley, and A. Harding, Predicting construction cost using multiple regression techniques. *Journal of construction engineering and management*, 2006. 132(7): p. 750-758.

[14] Abu Jamous, H., Parametric cost estimation of road projects using artificial neural networks. Parametric Cost Estimation of Road Projects Using Artificial Neural Networks, 2013.

[15] Skitmore, R.M. and S.T. Ng, Forecast models for actual construction time and cost. *Building and environment*, 2003. 38(8): p. 1075-1083.

[16] Hegazy, T. and A. Ayed, Neural network model for parametric cost estimation of highway projects. *Journal of Construction Engineering and Management*, 1998. 124(3): p. 210-218.

[17] Adeli, H. and M. Wu, Regularization neural network for construction cost estimation. *Journal of construction engineering and management*, 1998. 124(1): p. 18-24.

[18] Kim, G.-H., S.-H. An, and K.-I. Kang, Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. *Building and Environment*, 2004. 39(10): p. 1235-1242.

[19] Wilmot, C.G. and G. Cheng, Estimating future highway construction costs. *Journal of Construction Engineering and Management*, 2003. 129(3): p. 272-279.

[20] Oberlender, G.D. and S.M. Trost, predicting accuracy of early cost estimates based on estimate quality. *Journal of construction engineering and management*, 2001. 127(3): p. 173-182.

[21] Dukut, E.M., *The International Journal of Engineering and Science (IJES)* 2018. 2018, University of 17 August 1945 Surabaya. p. PP 41-49.