Conceptual Cost Estimate of Road Construction Projects in Saudi Arabia

Ibrahim Mahamid

Assistant Professor, Civil Engineering Department, Hail University, Hail, KSA.
E-Mail: imahamid@ymail.com

ABSTRACT

This study aims at developing regression models to predict the total construction cost of a road project in the early phases. The models were developed based on 52 sets of data collected in Saudi Arabia. The projects were implemented during 2011. The procedure used resulted in the formulation of 5 regression models; 2 of them include bid quantities of the major construction activities as independent variables and 3 include road length and road width as independent variables. The coefficient of determination (r^2) of the developed models is ranging from 0.65 to 0.97. This indicates that the relationship between the dependent and independent variables of the developed models is good and the predicted values from a forecast model fit with the real-life data. The value of the Mean Absolute Percentage Error (MAPE) of the developed regression models is tested. It is ranging from 17% to 42%. The results compare favorably with past researches which have shown that the estimate accuracy in the early stages of a project is between ±25% and ±50%. Such types of models are very useful, especially in simplicity and ability to be handled by calculator or using a simple computer program. It has a good benefit in estimating project cost at early stages of the project, since the information needed could be extracted easily from sketches or scope definition of the projects.

KEYWORDS: Conceptual estimation, Early stages, Cost estimation, Road construction, Regression, Saudi Arabia.

INTRODUCTION

Construction projects require high accuracy in cost estimation at the conceptual phase of a project. This is very important and crucial for proper project planning. At the early phases of a project, accurate cost advice is very important to the construction clients. It helps them to take a right decision regarding the feasibility of the proposed project. However, many difficulties arise when conducting cost estimation during the early phase. These difficulties could be: lack of conceptual information, lack of cost database, wrong cost estimation methods and many environmental, political, social and external uncertainties.

The term conceptual estimate is used to forecast the project's cost before completing the project design and specifications. The accuracy of an estimate is highly dependent on the available information about the proposed project at the time of estimate. Conceptual cost estimates are required at the early phase in which minimal information is available about the project, which makes the estimate less accurate. However, conceptual estimates are important to construction clients to examine before deciding to continue with a project. The early or conceptual phase is the first phase of a project in which the need is examined, alternatives are assessed, the goals and objectives of the project are established and a sponsor is identified (Holm et al., 2005). At this stage, the estimate accuracy is between ±25% and ±50% (Schexnayder and Mayo, 2003) due to

less defined project details. Regression analyses have been widely employed to develop such conceptual cost estimating models.

The objective of this research is to develop simple conceptual cost estimation models for contracted road construction projects that can be handled using a calculator or a simple computer program. To address that objective, mathematical models that describe the total project cost as a function of bid quantities of road construction activities and project size were developed. The estimation models were developed based on collected data for implemented road construction projects in Saudi Arabia. As these cost estimates are required at the early stage of the project, considerations were given to the fact that the input data for the required mathematical models could be easily extracted from sketches or scope definition of the projects.

LITERATURE REVIEW

The term "conceptual estimate" is used to describe the process of predicting a project's cost before the design of the project is completed (Sanders et al., 1992). The technique is used to estimate one characteristic of a system, usually its cost, from other physical and/or performance characteristics of the system (Ross, 2004).

Conceptual cost estimation is considered as the most significant starting process to influence the fate of a new project. The impact of cost variation on stakeholders decreases with the growing maturity of project planning and design. In practice, over- and under-estimation of project costs are both problematic (Sodikov, 2005).

Conceptual estimates are made for several reasons, including (Gould, 2005):
- Feasibility studies.
- Appropriation of funds.
- Appropriation of the project scope.

The type of estimate to be made and its accuracy depend upon many factors including the purpose of the estimates, how much is known about the project, and how much time and effort is spent in preparing the estimate (Humphreys, 2005).

Many of construction estimating models have been developed by previous research. The World Bank had developed an international database for road construction cost in developing countries; data were yielded in the form of Road Costs Knowledge System (ROCKS). It was designed to develop an international knowledge system on road work costs to obtain average and range unit costs based on historical data that could ultimately improve the reliability of new cost estimates. Data from 65 developing countries were used to make comparisons between estimated costs at appraisal and actual costs at completion (ROCKS, 2002).

Fragkakis et al. (2010) stated that the conceptual cost estimates rely on the conceptual design of the project and use only basic design technologies. Although they present the lowest expected accuracy, they are often used by key people involved in the construction process, thus playing a significant role. In their study, they described the development of prediction models for the material quantities of concrete and reinforcing and prestressing steel for three major bridge deck construction methods using regression analysis.

Petroutsatou et al. (2006) performed a study to develop early cost estimates for a road tunnel using multiple regression analysis. They gathered data for 33 projects implemented in Greece. The gathered data were used to establish correlation between geotechnical and construction parameters.

Sonmez (2004) conducted a study to develop conceptual cost estimates of building projects using regression analysis and neural networks. Three linear regression models were considered to identify the significant variables affecting the project cost. Two neural network models were developed to examine the possible need for nonlinear or interaction terms in the regression model.

Bell and Bozai (1987) developed multiple linear regression models for conceptual cost estimation for

road construction projects in Alabama, USA. A model that predicts the total project cost per mile as a function of bid quantities was developed. Hegazy and Ayed (1998) used a neural network approach to manage construction cost data and develop a parametric cost estimation model for highway projects.

Three cost prediction models were developed by Christian and Newton (1998) to determine accurate costs for road maintenance. These models were developed in the province of New Brunswick based on historical data during the period 1965-1994. Based on the models and the management review, it was concluded that maintenance funding needs to be increased by 25%.

Han et al. (2008) investigated the actual budgeting process in highway construction projects, under the research collaboration of the Korean Ministry of Construction and Transportation. Then, they developed the two-tiered cost estimation models of highway construction projects, considering the target goals for forecasting, allowable accuracy and available information level at each phase of a project's budgeting and initiation.

Mahamid (2011) conducted a study to describe the development of linear regression models to predict the costs of road construction activities based on 100 sets of data collected in Palestine. The procedure used resulted in the formulation of regression models which provide a means of determining the costs of road construction activities as a function of project's characteristics. The prediction models were developed for three major road construction activities which are earthworks, basecoarse works and asphalt works. Three groups of models for each activity were developed based on the used dependent variable; total cost of construction activity, cost per meter length and cost per meter square. The proposed independent variables were: road length, asphalt width, basecoarse width, terrain condition, soil drill ability and soil suitability.

Lowe et al. (2006) developed linear regression models to predict the construction cost of buildings, based on 286 sets of data collected in the United Kingdom. They identified 41 potential independent variables, and, through the regression process, showed five significant influencing variables: gross internal floor area (GIFA), function, duration, mechanical installations and piling.

Mahamid (2011) conducted a study that aims at developing early cost estimation models for road construction projects using multiple regression techniques, based on 131 sets of data collected in the West Bank in Palestine. 11 regression models were developed to estimate the total cost of road construction projects in US Dollar; 5 of them included bid quantities (e.g. earthworks, basecoarse, asphalt and furniture) and 6 included road length and road width as input variables.

RESEARCH METHOD

The research method is as follows:
1. Defining the problem.
2. Collecting the required data.
3. Selecting the forecasting method.
4. Selecting the model variables.
5. Developing the models.

Problem Definition

The first step in any engineering and scientific effort is to determine the objectives of the study or exercise. The objective of this study is to develop a mathematical model that describes the total cost of a road construction project as a function of the project's quantities and size.

Data Collection

Estimation requires extensive historical data. The data collected comprised of 52 road construction projects awarded during 2011 in Saudi Arabia. The data were tabulated to ensure that all costs were considered, that none is double-counted and that all are clearly defined. Considerations were taken to include almost equal numbers of small, medium and large projects in the collected data.

Selecting the Forecasting Method

There are many methods used to develop models to forecast the future construction cost. The estimation models include factors believed to influence construction costs. In general, the relationship between construction cost and the influencing factors is established based on past records of construction costs. Typically, the models established in this manner have been used to estimate the costs of individual contracts. These models with their relational structure are the only models expected to provide reliable long-term estimates (Wilmot and Cheng, 2003).

One of the widely used models is the regression estimation model. It is effective due to a well-defined mathematical approach, as well as because of being able to explain the significance of each variable and relationship between independent variables (Sodikov, 2005). The models developed in this study are of this type.

Proposed Variables

As the objective of this study is to develop simple estimation construction cost models in early stages of projects which could be easily handled using a calculator or a simple computer program. The models were developed based on quantitative factors that have significant impact on the contract price and that could be easily extracted from sketches at the early stage of the project. In this study, the following are the proposed variables:

- Total cost as a dependent variable.
- Independent variables;
  - X_1 = Earthwork; cut and fill (m^3);
  - X_2 = Base works (m^2);
  - X_3 = Asphalt works (m^2);
  - X_4 = Road length (m);
  - X_5 = Road width (m).

It should be also noticed that the flexible pavement of the road construction projects under this study consists of an asphalt layer (7 cm in thickness), a base layer (30 cm in thickness) in addition to earthworks.

Model Development

Once the variables to be included in the estimate equation have been identified, a series of mathematical models were developed using multiple regression analysis techniques. Regression models are intended to find the linear combination of independent variables which best correlates with dependent variables. The regression equation is expressed as follows:

$$Y = C + b_1X_1 + b_2X_2 + \ldots + b_nX_n \qquad (1)$$

where,

C: regression constant;

$b_1, b_2, \ldots, b_n$: regression estimates;

$X_1, X_2, \ldots, X_n$: independent variables;

Y: dependent variable.

The goal is to estimate the regression coefficients of Equation 1. Excel statistical tools were employed to perform regression analyses and to test the significance of the model. The models developed are discussed below.

RESULTS AND DISCUSSION

Once a set of probable predictors was identified for each project, mathematical models were developed using regression technique. Given the project quantities or project size for the predictor variables, the regression model predicts the total cost in US Dollars for that project.

Project Cost vs. Project Quantities

The above procedure resulted in the formulation of an equation that provides the total cost of a project using three predictor variables: earthworks, base works and asphalt works. The developed model is as follows:

$$\text{Total cost (\$)} = 11256.65 + 3.23X_1 + 2.15X_2 + 4.18X_3$$
$$(\text{model 1})$$

where,

X_1 = Earthwork; cut and fill (m^3);

X_2 = Base works (m^2);

X_3 = Asphalt works (m^2).

The coefficient of multiple determination (R^2) for the developed model is 0.91, indicating a good relationship between dependent and independent variables. The regression statistics for the developed model are shown in Table 1. The analysis of variance test confirmed the statistical significance of the model at a significance level of 0.05 (P-value less than 0.05).

Table 1. Project cost ($) as a function of project quantities

Regression Statistics		Variables	Coef.	P-value
Multiple R	0.93	Intercept	11256.65	0.0015
R^2	0.91	Earthworks (m^3)	3.23	1.23E-06
Adjusted R^2	0.80	Base works (m^2)	2.15	$\cong 0$
Observations	52	Asphalt (m^2)	4.18	$\cong 0$
F	107.2			

Three individual regression models are developed. The models describe the costs of the major activities in road construction: earthworks, base works and asphalt works as a function of their bid quantities. The results are shown in Table 2. The analysis of variance test confirmed the statistical significance of the models at a significance level of 0.05.

Table 2. Cost models of major road construction activities *vs*. activities' quantities

Activity	Variables	Coef.	r^2	P-value	F	model
Earthworks	Earthwork quantity (m^3) = X_1	3.11	0.71	0.004	117.3	Earthwork cost ($) = $3.11X_1$
Base works	Base quantity (m^2) = X_2	1.54	0.92	≈ 0	4623.8	Base cost ($) = $1.54X_2$
Asphalt works	Asphalt quantity (m^2) = X_3	3.92	0.97	≈ 0	5126.4	Asphalt cost ($) = $3.12X_3$

These models help in estimating the costs of the mentioned major activities in road construction individually. The statistics also show that the cost of the above-mentioned major activities in road construction constitutes about 80% of the total project cost, and so it is possible to use their individual cost estimation models to predict the total project cost.

The correlation between the input variables is tested; the results are shown in Table 3. The results of r^2 show that there is a high correlation between asphalt works and base works ($r^2 = 0.95$) and a good correlation between asphalt works and earthworks ($r^2 = 0.78$).

Table 3. Correlation between input variables

Correlation between:	r^2
Asphalt works – base works	0.95
Asphalt works – earthworks	0.78

As a result of the correlation between input variables, a regression model describing the total project cost as a function of asphalt quantity is developed. The regression statistical results are shown in Table 4. The model is useful in estimating project, cost at early stages of the project, since the information needed is only the asphalt quantity, and so the estimation could be achieved within minutes. The developed model is:

Total project cost (\$) = 17643.25 + 8.53X_3 (model 2)   where, X_3 = asphalt quantity (m^2).

Table 4. Multiple regression results among total project cost and pavement quantity

Regression Statistics		Variables	Coef.	P-value
Multiple R	0.91	Intercept	17643.25	0.0002
R^2	0.88	Asphalt (m^2)	8.53	≈ 0
Observations	52			
F	1163.1			

Cost as a Function of Project Size

A regression model that describes the total cost of a road construction project as a function of road width and road length is developed. The coefficient of multiple determination (R^2) for the developed equation is 0.95. The regression statistical results for the developed model are shown in Table 5. The developed model is as follows:

Total cost (\$) = 107.5X_4 + 7629.72X_5 (model 3)
  where,
X_4 = Road length (m);
X_5 = Road width (m).

Table 5. Multiple regression results between total project cost and road length and width

Regression Statistics		Variable	Coefficient	t Stat.	P-value	F
Multiple R	0.98	road length	107.50	7.14	0.00	81.32
R^z	0.95	road width	7629.72	2.08	0.07	
Adjusted R Square	0.82					
Observations	52					

Table 5 shows that the p-value of the road width variable is 0.07 which is higher than 0.05, meaning that it is not significant to be included in the model. As a result, a model uses cost/lane as a dependent variable, and road length as the only input variable is formulated. The regression statistical results for the developed model are shown in Table 6. The formulated equation is:

Cost (\$)/lane = 45.58X_4 (model 4)

Table 6. Multiple regression results between cost/lane and road length

Regression Statistics		Variable	Coefficient	t Stat.	P-value	F
Multiple R	0.96	road length	45.58	36.75	0.00	1415.65
R^2	0.94					
Adjusted R^2	0.87					
Observations	52					

where X_4 = road length (m).

A model using variables' interaction has also been developed. The developed equation is:

Total cost (\$) = 13.48 $X_4 X_5$.   (model 5)

The coefficient of determination (r^2) for the developed equation = 0.97. The regression statistical

results for the developed model are shown in Table 7. The results show that the analysis of variance test confirmed the statistical significance of the model at a significance level of 0.05.

Table 7. Multiple regression results betenn total project cost and road size

Regression Statistics		Variable	Coefficient	t Stat.	P-value	F
Multiple R	0.98	road size	13.48	40.37	0.00	1629.39
R^2	0.97					
Adjusted R^2	0.88					
Observations	52					

Equations that describe the cost of the major road construction activities as a function of road length and road width are developed. The best fit models are achieved when variables interaction is used and the intercept value = 0. The results are shown in Table 8.

The table shows that r^2 values are high; this indicates a good relationship between dependent and independent variables. These models help in cost estimation of the major road construction activities individually.

Table 8. Cost of major road construction activities ($) as a function of project size (m^2)

Activity	Coef.	r^2	P-value	F-test	Regression equation
Earthworks	2.31	0.61	0.002	97.35	$2.31X_4X_5$
Base works	2.46	0.91	≈0	1035.69	$2.46X_4X_5$
Asphalt works	4.12	0.96	≈0	3124.73	$4.12X_4X_5$

Testing the Accuracy of the Developed Models

The mean absolute percentage error (MAPE) is used to measure the accuracy of the developed models. The following formula is used to compute the MAPE (Lowe et al., 2006):

$$MAPE = (1/n) \sum_{i=1}^{n} \left| (A_i - F_i)/ A_i \right| \qquad (2)$$

where,

A_i is the actual value;

F_i is the forecast value;

n is number of fitted points.

Table 9 shows a summary of the tested regression

models. 5 models are tested; two of them include bid quantities as independent variables (models 1 and model 2), while the other three models include road length and road width as independent variables (models 3 through 5).

It should be noticed that in the very early stages the bill of quantity (BOQ) is not available, meaning that the models using road width and length (models 3 through 5) are more easy and fit to be used in this stage of a project. Later, when the BOQ is available, the models based on BOQ (models 1 and model 2) may be used.

Table 9. Summary of the developed regression models

Model No.	Regression models
1	Total cost (\$) = 11256.65 + 3.23X_1 + 2.15X_2 + 4.18X_3
2	Total project cost (\$) = 17643.25 + 8.53X_3
3	Total cost (\$) = 107.5X_4 + 7629.72X_5
4	cost (\$)/lane = 45.58X_4
5	Total cost (\$) = 13.46 X_4X_5

where,

X_1 = Earthworks; cut and fill (m^3);

X_2 = Base works quantity (m^2);

X_3 = Pavement quantity (m^2);

X_4 = Road length (m);

X_5 = Road width (m).

Table 10 shows the MAPE resulting from using the developed regression models to estimate the total cost of 52 data sets of road construction projects. The results show the following:

- The MAPE of the developed models is ranging from 17% to 42% which compare favorably with past researches which have shown that the estimate accuracy in the early stages of a project is between ±25% and ±50% (Lowe et al., 2006; Schexnayder et al., 2003; Sodikov, 2005).

- It can be seen that the models that used bid quantities as independent variables are more accurate than the models that used road length and road width as independent variables.

- The table shows that for the models that used bid quantities as independent variables, when the number of work items involved in the model increases, the r^2 value increases and MAPE value decreases.

- The table shows that for the models that used road length and width as independent variables, the MAPE is ranging from 19.6% to 41.6%. It can be seen that the most accurate model is the model that

used variables' interaction as independent variable (model 5), followed by model 3 and model 4, respectively.

Table 10. Mean absolute percentage error (MAPE) and r^2 of the developed regression models

Model No.	MAPE	r^2
1	17.2%	0.91
2	19.4%	0.88
3	25.3%	0.95
4	32.1%	0.94
5	19.6%	0.97

CONCLUSIONS

This study aims at developing conceptual cost estimation models for road construction projects using multiple regression techniques. The models were developed based on 52 sets of data collected in Saudi Arabia. Such types of models are very useful, especially in simplicity and ability to be handled by a calculator or a simple computer program. They have good benefits in estimating project cost at early stages of the project, since the information needed could be extracted easily from sketches or scope definition of the projects.

It must be remembered that an estimated project cost is not an exact number, but it is an opinion of probable cost. The accuracy and reliability of an estimate are totally dependent upon how well the project scope is defined and the time and effort expended in the preparation of the estimate.

In this study, 5 regression models are developed; 2 of them include bid quantities as independent variable and 3 include road length and width. The coefficient of determination (r^2) of the developed models is ranging from 0.65 to 0.97. This indicates that the relationship between the dependent and independent variables of the developed models is good and the predicted values from a forecast model fit with real-life data.

The values of the mean absolute percentage error (MAPE) of the developed regression models are ranging from 17% to 42%. The results compare favorably with previous research which has shown that the estimate accuracy in the early stages of a project is between ±25% and ±50%. The findings reveal that the models that used bid quantities as independent variable are more accurate than those that used road length and width as independent variable, but they require more information.

Based on the results, the following recommendations are suggested to improve the cost estimation process in Saudi Arabia:

- Continuous updating on material prices, equipment rates and labor rates by the Government or industry associations.
- Developing a computerized database that enables estimators and researchers to use it easily in their estimates and studies.
- The models should be revised periodically using up-to-date data.
- Further studies should be conducted to include more construction variables as independent variables in the cost estimation models.

REFERENCES

Bell, L. and Bozai, G. 1987. Conceptual cost estimating for highway construction projects. *AACE Transactions,* C.6.1-C.6.4.

Christian, J. and Newton, L. 1998. Highway construction and maintenance costs. Construction Engineering and Management Group, University of New Brunswick, Fredericton, NB E3B 5A3, Canada.

Fragkakis, A., Lambropoulos, S. and Pantouvakis, J. 2010. A cost estimate method for bridge superstructures using regression analysis and bootstrap. *Organization, Technology and Management in Construction Journal,* 2 (2): 182-190.

Gould, F. 2005. Managing the construction process: estimating, scheduling and project control. Pearson Education, Inc., New Jersey.

Han, S., Kim, D. and Kim, H. 2008. Two-staged early cost estimation for highway construction projects. Report, Civil and Environmental Engineering, Yonsei University, Seoul, Korea.

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

Holm, L., Schaufelberger, J., Griffin, D. and Cole, T. 2005. Construction cost estimating: process and practices. Pearson Education, Inc., New Jersey.

Humphreys, K. 2005. Project and cost engineer's handbook (4th edition). Marcel Dekker, USA.

Lowe, D., Emsley, M. and Harding, A. 2006. Predicting construction cost using multiple regression techniques. *Journal of Construction Engineering and Management,* 132 (7): 750-758.

Mahamid, I. 2011. Cost and time overrun in road construction: magnitude, causes and cost estimating models. PhD Thesis, NTNU, Trondheim, Norway.

Mahamid, I. 2011. Early cost estimating for road construction projects using multiple regression techniques. *Australasian Journal of Construction Economics and Building,* 11 (4): 78-101.

Petroutsatou, C., Lambropoulos, S. and Pantouvakis, J. 2006. Road tunnel early cost estimates using multiple regression analysis. *Operational Research Journal*, 6 (3): 311-322.

ROCKS. 2002. Road Costs Knowledge System (v. 2.01), (The World Bank, Transport Unit – TUDTR). Retrieved August 4, 2011, from http://www.worldbank.org/transport/roads/rd_tools/rocks_main.html.

Ross, S. 2004. Introduction to probability and statistics for engineers and scientists (3rd edition). Elsevier, Amsterdam.

Sanders, S., Maxwell, R. and Glagola, C. 1992. Conceptual estimating models for infrastructure projects. *Cost Engineering,* 34 (8).

Schexnayder, C. and Mayo, R. 2003. Construction management fundamentals. McGraw-Hill Higher Education, Boston, MA.

Sodikov, J. 2005. Cost estimation of highway projects in developing country: artificial neural network approach. Graduate Project, Department of Civil and Environmental Engineering, Saitama University.

Somenz, R. 2004. Conceptual cost estimation of building projects with regression analysis and neural networks. *Canadian Journal of Civil Engineering*, 31(4): 677-683.

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