



Early Cost Estimating for Road Construction Projects Using Multiple Regression Techniques

Ibrahim Mahamid, (Hail University, Saudi Arabia)

Abstract

The objective of this study is to develop early cost estimating models for road construction projects using multiple regression techniques, based on 131 sets of data collected in the West Bank in Palestine. As the cost estimates are required at early stages of a project, considerations were given to the fact that the input data for the required regression model could be easily extracted from sketches or scope definition of the project. 11 regression models are developed to estimate the total cost of road construction project in US dollar; 5 of them include bid quantities as input variables and 6 include road length and road width. The coefficient of determination r^2 for the developed models is ranging from 0.92 to 0.98 which indicate that the predicted values from a forecast models fit with the real-life data. The values of the mean absolute percentage error (MAPE) of the developed regression models are ranging from 13% to 31%, the results compare favorably with past research which have shown that the estimate accuracy in the early stages of a project is between $\pm 25\%$ and $\pm 50\%$.

Keywords: Cost estimating, Regression, Road construction, Early estimate

Introduction

The term early 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 (Rose, 1982). This technique involves life cycle costs, a detailed data base, and the application of multivariable correlation (Black, 1984).

Early cost estimating is considered as the most significant starting process to influence the fate of a new project (Sodikov, 2005). The accuracy of cost estimation improves toward the end of the project due to detailed and precise information. 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, estimate accuracy is between $\pm 25\%$ and $\pm 50\%$ (Schexnayder et al., 2003) due to less defined project details.

Cost estimation of construction projects with high accuracy at the early phase of project development is crucial for planning and feasibility studies. Construction clients require early and accurate cost advice prior to site acquisition and commitment to build in order to enable them to take a right decision regarding the feasibility of proposed project. However, a number of difficulties arise when conducting cost estimation during the early phase. Major problems include lack of preliminary information, lack of database of works costs, lack of appropriate cost estimation methods, and the involvement of many environmental, political, social and external uncertainties. Given its significance, conventional tools such as regression analysis have been widely employed to tackle the problem.

In Palestine, the construction industry is one of the main economic driving sectors, supporting the Palestinian national economy. It contributes to 26% of the Palestinian GDP. It also plays a basic role in providing homes, public facilities and infrastructure, absorbing

work forces and improving the whole Palestinian national economy. However, many construction projects report poor cost, time and quality performances. According to Mahamid et al. (2010), the cost overrun is a phenomenon in road construction projects in Palestine. Through a field study that investigated the cost diverge in road construction based on data of 169 road construction projects, they concluded that 100% of the projects suffer from cost diverge. They found that among these 169 projects, 76.33% were underestimated and 23.67 were overestimated. Therefore, it is very important issue to improve accurate models that help in reducing cost diverge in construction projects.

This study presents regression models that describe the total cost of road construction project as a function of bill of quantity (BOQ) and project size (i.e. road length and width). The estimating models were developed based on collected data for 131 awarded road construction projects in the West Bank in Palestine. As these cost estimates are required at early stages of the project, considerations were given to the fact that the input data for the required model could be easily extracted from sketches or scope definition of the project.

Objectives

The objectives of this study are:

- To develop preliminary cost estimating models of road construction projects as a function of unit quantities of road construction activities
- To develop early cost estimating models of road construction projects as a function of project size (i.e. road length and road width)

Literature Review

Ahuja et al. (1994) state that estimating is the primary function of the construction industry; the accuracy of cost estimates starting from an early phase of a project through the tender estimate can affect the success or failure of a construction project. They also state that many failures of construction projects are caused by inaccurate estimates.

A cost estimate establishes the base line of the project cost at different stages of development of the project. As Hendrickson et al. (1989) point out, a cost estimate at a given stage of project development represents a prediction provided by the cost engineer or estimator on the basis of available data. Gould (2005) defined estimate as an appraisal, an opinion, or an approximation as to the cost of a project prior to its actual construction. According to Jelen et al. (1983), estimating is the heart of the cost engineer's work and consequently it has received appropriate attention over the years.

Holm et al. (2005) lists several reasons for making estimate, including:

- Feasibility studies
- Selection from alternate design
- Selection from alternate investment
- Appropriation of funds
- Presentation of bids and tenders

A number of cost prediction models have been developed. Bell (1987) developed multiple linear regression models for preliminary cost estimating to be used by Alabama Highway Department (AHD) for long range cost forecasting. The total project cost per mile is the function of a list of probable predictors comprising line items, such as quantities of work items per mile.

Mahamid et al. (2010) developed multiple linear regression models for preliminary cost estimating of road construction activities as a function of project's physical characteristics such as terrain conditions, ground conditions and soil drillability.

Hegazy et al., (1998) used a neural network approach to manage construction cost data and develop a parametric cost estimating model for highway projects. They introduced two alternative techniques to train network's weights: simplex optimization (Excel's inherent solver function), and GAs (genetic algorithms).

Three cost prediction models were developed by Christian and Newton (1998) in order to determine an accurate cost 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 needed to be increased by 25%.

Lowe et al. (2006) developed linear regression models in order 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 such as gross internal floor area (GIFA), function, duration, mechanical installations, and piling.

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. They then developed two-tiered cost estimation models of highway construction projects, considering the target goals for forecasting, allowable accuracy, and available information at each phase of a project budgeting and initiation.

Recent work reveals that there are still many problems in cost estimation at the conceptual stage of a project cycle. The World Bank had developed an international database for road construction cost in developing countries; the data was yielded in form of ROad Costs Knowledge System (ROCKS). It was designed to develop an international knowledge system of road work costs to obtain average and range unit costs based on historical data that could ultimately improve the reliability of new cost estimates. In that study, data from 65 developing countries were used to make comparisons between estimated costs at appraisal and actual cost at completion. Among these projects 62% were overestimated, and the rest underestimated (ROCKS, 2002).

Research Methodology

The methodology approach in this research is shown in figure 1, each step will be discussed in succeeding paragraphs.

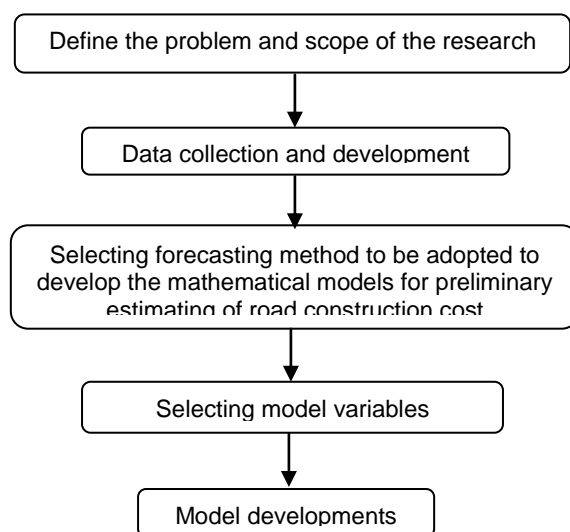


Figure 1 Research methodology

Problem Definition

The objective of this study is to develop regression models that estimate the total cost of road construction projects as a function of bid quantities and project size.

Data Collection

The estimating technique requires an extensive historical data base. The data were collected from contracts awarded by Palestinian agencies, the clients for road construction projects, awarded in the West Bank. The data collected comprised 131 projects awarded over the years 2004-2008. The data were tabulated to ensure that all costs were considered, none is double-counted and all are clearly defined.

All the data were deflated to year 2008. The construction cost index of 2008 obtained from Palestinian Economic Council for Development and Reconstruction (PECDAR) is used to deflate the data. Table 1 shows the index values over the years 2004-2008, the base year is 2008 (index = 1).

Year	2004	2005	2006	2007	2008
Cost index	0.60	0.62	0.66	0.88	1

Table 1 Construction cost index value in the West Bank (PECDAR, 2009)

To deflate the cost of a project in a certain year to cost in year 2008, its cost is divided on the cost index value of that year.

The following guidelines were taken into consideration during the data collection:

1. Distribution among year of award

A consideration is taken to have approximately equal number of awarded projects over the years 2005-2008. In 2004, only projects that awarded in the last quarter of the year were considered in the study in order to avoid the Second Intifada effects. Table 2 shows the distribution of projects based on year of award in the collected data.

Year of award	No. of projects
2004	9
2005	34
2006	34
2007	28
2008	26

Table 2 Projects distribution based on year of award in the collected data

2. Project size

Road construction projects in the West Bank are classified by PECDAR based on their cost as shown in Table 3.

Category	Classification
< USD150000	small project
USD150000 - USD300000	medium project
> USD300000	large project

Table 3 Classification of road construction projects in the West Bank (PECDAR, 2009)

In this study, a consideration is taken to have approximately equal number of projects under each category. Table 4 shows the projects distribution based on project cost.

Category	No. of projects
small	46
medium	43
large	42

Table 4 Projects distribution based on project cost in the collected data

Moreover, a consideration is taken to have approximately equal number of projects under each cost category per each year of award. Table 5 shows the projects distribution based on project cost per each year of award.

Year	small	medium	large
2004	1	6	2
2005	10	10	14
2006	10	14	10
2007	14	6	8
2008	11	7	8

Table 5 Projects distribution based on size category in each year of award

Selection of Forecasting Method

Previous studies show many methods which are used to forecast future construction costs. Some of these models describe construction cost as a function of factors believed to influence construction cost. The relationship between construction cost and these factors have been established from past records of construction cost. Typically, the models established in this manner have been used to estimate the cost of individual contracts. These models with their relational structure are the only models expected to provide reliable long-term estimates (Wilmot and Cheng, 2003).

Regression estimating models are widely used in cost estimation. They are effective due to a well-defined mathematical approach, as well as being able to explain the significance of each variable and the relationships between independent variables (Sodikov, 2005). The developed models in this study are of this type.

Proposed Model's Variables

Many qualitative and quantitative factors affect the contract price as shown in previous studies. It is striking, however, that most of past construction models used only a few of them because the lack of information available in early stages of a project and information about qualitative factors surrounding each project are difficult to obtain.

As the objective of this study is to develop estimating models that can be handled easily using calculators or simple computer programs in the early stages of project, the models were developed based on quantitative factors that have significant impact on contract price and that could be easily extracted from sketches at an early stages of a project. The following variables are used:

- Total project cost as output (dependent) variable which has been represented in year 2008 US dollar; it ranges from \$42000 - \$2000000
- Input (independent) variables are shown in Table 6.

Models Development

Once the variables to be included in the estimate equation have been identified, a series of 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:

$$Y = C + b_1X_1 + b_2X_2 + \dots + b_nX_n \quad (1)$$

Where,

C: regression constant

b_1, b_2, \dots, b_n : regression estimates

X_1, X_2, \dots, X_n : independent variables

Y: dependent variable.

Where all variables are significant at the 0.05 level. The variables are described in Table 6.

Description	Unit	Range
Earthwork; cut, fill, and topping	m ³	250 - 28000
Basecourse	m ³	350 - 22500
Pavement	m ²	1000 - 87000
Curb Stone	ml**	200 - 7800
Retaining wall concrete	m ³	20 - 450
Side walk concrete	m ³	5 - 570
Road marking	ml**	100 - 6500
Road length	m	700 - 14000
Road width	m	3 - 10

Table 6 Input variables description in the collected data set

*asphalt layer thickness is 5 cm, **ml: meter length

Results and Discussion

Cost as a Function of Bid Quantities

Once a set of probable predictors were identified for each project, a statistical model was developed using multiple regression technique. Given the quantities per project for the predictor variables, the regression model predicted the total cost in US dollars for that project, based on six statistically significant variables as shown in Table 7, where the p-values for all coefficients considered in the model are less than or equal to 0.05:

$$\begin{aligned} \text{Total cost (\$)} = & 16773,8 + 4,6X_1 + 32,5X_2 + 14,1X_3 + 156,5 X_5 \\ & + 485,5X_6 - 15,5X_7 \end{aligned} \quad (\text{model 1})$$

Regression Statistics		Variables	Coef.	P-value
Multiple R	0,99	Intercept	16773,8	0,0063
R Square	0,98	Earthworks (m ³)	4,6	0,0103
Adjusted R Square	0,98	Basecourse (m ³)	32,5	4E-12
Observations	131	Pavement (m ²)	14,1	2E-30
F	1031,5	Retaining wall conc. (m ³)	156,5	0,0632
		Side walk conc. (m ³)	485,5	1E-10
		Road marking (ml)	-15,5	0,0537

Table 7 Multiple regression results among total project cost and activities' quantities

Data analyses show that the earthworks, base works, asphalt works and furniture works (i.e. all road construction activities excluding earthworks, base works and asphalt works) are the major activities in the 131 examined road construction projects. Figure 2 shows the average contribution of these activities in the total project cost.

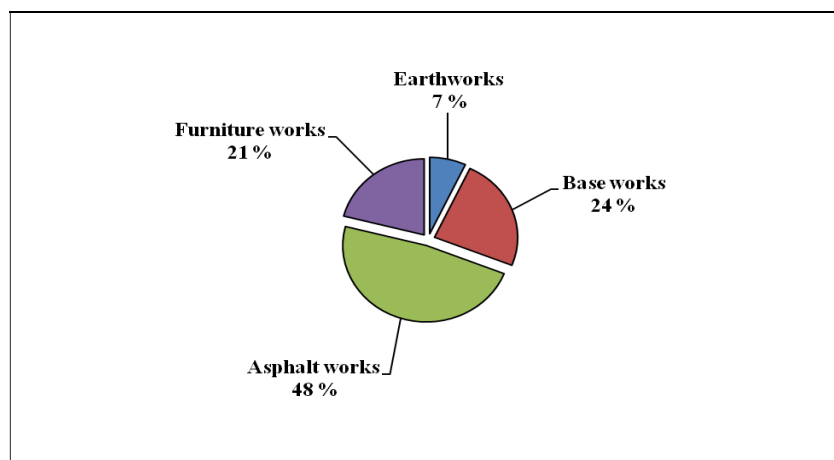


Figure 2 Average contribution of major road construction activities in the total project cost

As the furniture works include a lot of work items, an equation that describes the total project cost as a function of quantities of major construction activities excluding furniture works is formulated; this makes the model more easy to be used in the early stages where a little information is available. The coefficient of determination r^2 of the developed model is 0.97. The analysis of variance test confirmed the statistical significance of the model at a significance level of 0.05 as shown in Table 8. The developed model is:

$$\text{Total cost (\$)} = 23241 + 8,5 X1 + 28 X2 + 15,2X3 \quad (\text{model 2})$$

Where,

- X1 = Earthwork; cut, fill, and topping quantities (m^3)
- X2 = Basecourse quantity (m^3)
- X3 = Pavement quantity (m^2)

Regression Statistics			Variables	Coef.	P-value
Multiple R	0,98		Intercept	23241	0,0013
R Square	0,97		Earthworks (m^3)	8,5	1,43E-05
Adjusted R Square	0,97		Basecourse (m^3)	28,0	1,79E-07
Observations	131		Pavement (m^2)	15,2	8,26E-28
F	1376,2				

Table 8 Multiple regression results among total project cost and major activities quantities

Another way to estimate the total project cost is to find the estimated cost for each construction activity and then to sum them up. The formulated model will be in the form of:

$$\text{Total cost (\$)} = \text{Earthworks cost} + \text{Basecourse works cost} + \text{Asphalt works cost} + \text{Furniture works cost}$$

Three individual regression models that describe the cost of earthworks, basecourse works and asphalt works as a function of bid quantities are developed. The results are shown in Table 9. The analysis of variance test confirmed the statistical significance of the models at a significance level of 0.05.

Activity	Variables	Coef.	r ²	P-value	F	Best fit model
Asphalt works	Pavement quantity (m ²) = X3	12,19	0,97	≈ 0	4503,8	Asphalt cost(\$) = 12,19X1
Base works	Basecourse quantity (m ³) = X2	23,00	0,97	≈ 0	5135,7	Basecourse cost(\$)= 23X2
Earthworks	Earthwork quantity (m ³) = X1	6,00	0,89	≈ 0	1032,9	Earthwork cost(\$) = 6X3

Table 9 Regression models among cost of each major activity in road construction and its quantity

The furniture cost can be estimated using one of the following equations (Table 10):

Equation #	Variables	P-value	F	r ²	MAPE	Best fit model
2	curb stone (ml) = X4	1,89E-11	118	0,81	72,40 %	Furniture cost(\$) = 37,4X4 + 324,6X5 + 388,5X6 + 19,3X7
	retaining wall conc. (m ³) = X5	0,000342				
	side walk concrete (m ³) = X6	2,55E-07				
	road marking (ml) = X7	0,013973				
3	road size (m ²) = X8*X9	5,62E-49	560,5	0,81	48 %	Furniture cost(\$) = 4,8X8*X9
4	pavement (m ²) = X3	7,80E-58	814,5	0,86	32 %	Furniture cost(\$) = 5,6X3

Table 10 Cost estimating models for furniture works

Table 10 shows that Equation 3 is the best to be used for furniture cost estimation as it has the highest r² and lowest MAPE. The coefficient of determination r² for model 4, using Equation 4 for furniture cost estimation (shown in Table 11), is 0.97. Then the developed model is:

$$\begin{aligned} \text{Total project cost (\$)} &= 6X1 + 23X2 + 12,19X3 + [5,6X3] \\ \Rightarrow \text{Total project cost (\$)} &= 6X1 + 23X2 + 17,79X3 \end{aligned} \quad (\text{model 3})$$

Model 3 is similar to model 2, but model 3 could be used to estimate cost of one individual construction activity (i.e. asphalt cost, basecourse cost, earthworks cost, and furniture cost). It can be seen that model 2 is better for estimates of the total project cost because it has better accuracy (MAPE) than model 3 as shown later in Table 20.

The correlation among input variables is tested; the results are shown in Table 11. The results of r² show that there is a high correlation between pavement quantity and basecourse quantity, a medium correlation between pavement quantity and earthworks quantity, and a low correlation between pavement quantity and other input variables.

correlation among quantities of:	r ²
pavement - basecourse	0,88
pavement - earthworks	0,64
pavement - retaining wall concrete	0,32
pavement - curbstone	0,47
pavement - road marking	0,29
pavement - side walk concrete	0,21

Table 11 Correlation among used input variables

As a result of the correlation between input variables, a regression model describing the total project cost as a function of pavement quantity is developed. The regression statistics results are shown in Table 12. The model is useful in estimating project cost at early stages of the project since the information needed is only the pavement quantity, and so the estimation could be achieved within minutes. The developed model is:

$$\text{Total project cost (\$)} = 34832,5 + 23,7X3 \quad (\text{model 4})$$

Where, X3 is the project's pavement quantity (m²).

Regression Statistics		Variables	Coef.	P-value
Multiple R	0,97	Intercept	34832,5	0,0001
R Square	0,95	pavement (m ²)	23,7	5,9E-86
Observations	131			
F	2466,7			

Table 12 Multiple regression results among total project cost and pavement quantity

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 determination r² for the developed equation is 0.93. The regression statistics results for the developed model are shown in Table 13. The developed model is:

$$\text{Total cost (\$)} = 142,6X8 + 28882,7X9 - 196877,7 \quad (\text{model 5})$$

Where,

X8 = Road length (m)

X9 = Road width (m)

Regression Statistics		Variables	Coef.	P-value
Multiple R	0,97	Intercept	-196876,7	3,78E-10
R Square	0,93	Road width (m)	28882,7	9,43E-09
Adjusted R Square	0,93	Road length (m)	142574,6	1,67E-75
Observations	131			
F	920,5			

Table 13 Multiple regression results among total project cost and road length and width

It can be seen that the model has a large constant (-196877,7) and so in small projects, the estimated total cost using the model will be not realistic or it may even be negative e.g. for a road width of 5 m and road length of 370 m, the estimated cost is 0. It may give a realistic estimated value when road length is more than 1 km and road width is more than 5 m.

Therefore, a trial was performed to develop a model with zero intercept value, the results shown in Table 14. The developed model is:

$$\text{Total cost (\$)} = 139,3X8 - 542,1X9 \quad (\text{model 6})$$

Regression Statistics		Variables	Coef.	P-value
Multiple R	0,98	Intercept	0	#N/A
R Square	0,95	Road width(m)	-542,1	0,80
Adjusted R Square	0,94	Road length (m)	139,3	1,78E-67
Observations	131			
F	1305,2			

Table 14 Multiple regression results among total project cost and road length and width when intercept = 0

Table 14 shows that the p-value of the road width variable is 0.8 which is higher than 0.05, meaning it is not significant to be included in the model. As a result, the model uses road length as the only input variable is formulated. The regression statistics results for the developed model are shown in Table 15. The formulated equation is:

$$\text{Total cost (\$)} = 145,9X8 - 32626 \quad (\text{model 7})$$

Regression Statistics		Variables	Coef.	P-value
Multiple R	0,96	Intercept	-32626	0,01154
R Square	0,92	Road length (m)	145847	3,5E-71
Adjusted R Square	0,92			
Observations	131			
F	1403,4			

Table 15 Multiple regression results among total project cost and road length

The developed model has high r² and low p-value (less than 0.05), but for very small projects, it will give unrealistic results; e.g. at road length 220 m, the estimated cost is 0. The model may give good estimated results when the road length is higher than 600 m. Therefore, a model is developed using the road length only as independent variables with zero intercept value, the results are shown in Table 16. The developed model is:

$$\text{Total cost (\$)} = 138,53X8 \quad (\text{model 8})$$

Regression Statistics		Variables	Coef.	P-value
Multiple R	0,98	Road length (m)	138,53	4,07E-88
R Square	0,95			
Adjusted R Square	0,95			
Observations	131			
F	2629,25			

Table 16 Multiple regression results among total project cost and road length when intercept = 0

A model using variables interaction has also been developed, the resulting equation is:

$$\text{Total cost (\$)} = 20,73 X8 * X9 \quad (\text{model 9})$$

The coefficient of determination r^2 for the developed equation = 0.96. The regression statistics results for the developed model are shown in Table 17. The results show that the analysis of variance test confirmed the statistical significance of the model at a significance level of 0.05.

Regression Statistics		Variables	Coef.	P-value
Multiple R	0,98	Intercept	0	#N/A
R Square	0,96	Road size (m ²)	20,73	1,69881E-89
Adjusted R Square	0,95			
Observations	131			
F	2767,4			

Table 17 Multiple regression results among total project cost and road size

Equations that describe the cost of the major road construction activities as a function of road length and width are developed. The best fit models are achieved when variables interaction is used and intercept value = 0. The results are shown in Table 18. The Table shows that r^2 values are high. This indicates a good relationship between dependent and independent variables.

Activity	Coef.	r ²	P value	F-test	Regression equation
Earthworks	1,56	0,78	≈0	454,42	1,56X8X9
Basecourse works	4,60	0,98	≈0	5067,75	4,6X8X9
Asphalt works	10,00	0,93	≈0	1780,67	10X8X9
Furniture works	4,58	0,81	≈0	560,65	4,58X8X9

Table 18 Cost of major road construction activities (\$) as a function of project size (m²)

Then the total project cost can be calculated by summing up the cost of construction activities (model 10). The coefficient of determination r^2 for the developed model is 0.92.

$$Total\ cost\ (\$) = Earthworks\ cost + Basecourse\ works\ cost + Asphalt\ works\ cost + Furniture\ cost \quad (model\ 10)$$

Where;

$$\begin{aligned} Earthworks\ cost &= 1,56X8X9 \\ Base\ works\ cost &= 4,6\ X8X9 \\ Asphalt\ works\ cost &= 10\ X8X9 \\ Furniture\ works\ cost &= 4,58\ X8X9 \end{aligned}$$

It can be seen that model 10 is similar to model 9 (both use variables interaction), but model 10 could be used to estimate the cost of one individual construction activity (i.e. asphalt cost, basecourse cost, earthworks cost, and furniture cost). Also it has better accuracy (MAPE) than model 9 as shown in Table 20.

Testing 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):

$$MAPE = (1/n) * \sum_{i=1}^n | (A_i - F_i) / A_i | \quad (5)$$

Where,

A_i is the actual value
 F_i is the forecast value
 n is number of fitted points

Table 20 shows a summary of the developed regression models in the study, 10 models are developed to estimate the total cost of road construction project in US dollar; 4 of them include bid quantities as independent variables (models 1 through 4), while the other 6 models include road length and road width as independent variables (models 5 through 10).

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 5 through 10) are more easy and fit to be used. Later, when the BOQ is available, the models based on BOQ (models 1 through 4) may be used.

Model No.	Regression models
1	Total cost (\$) = 16773,8 + 4,6X1+ 32,5X2 + 14,1X3 + 156,5 X5 + 485,5X6 -15,5X7
2	Total cost (\$) = 23241 + 8,5 X1 + 28 X2 + 15,2X3
3	Total cost (\$) = 6X1+ 23X2 + 12,19X3 + Furniture cost; where, furniture cost = 5,6X3
4	Total cost (\$) = 34832,5 + 23,7X3
5	Total cost (\$) = 142,57X8 + 28882,7X9– 196877,7
6	Total cost (\$) = 139,3X8 – 542,1X9
7	Total cost (\$) = 145,9X8- 32626
8	Total cost (\$) = 138,53X8
9	Total cost (\$) = 20,73 X8 X9
10	Total cost (\$) = Earthworks cost + Base works cost + Asphalt works cost + Furniture cost Where; Earthworks cost = 1,56X8X9, Base works cost = 4,6 X8X9, Asphalt works cost = 10 X8X9, Furniture works cost = 4,58 X8X9.

Table19 Summary of the developed regression models

Where,

- X1 = Earthwork; cut, fill, and topping quantities (m3)
- X2 = Basecourse quantity (m3)
- X3 = Asphalt quantity (m2)
- X4 = Curb stone quantity (ml)
- X5 = Retaining walls concrete quantity (m3)
- X6 = Side walk concrete quantity (m3)
- X7 = Road marking (ml)
- X8 = Road length (m)
- X9 = Road width (m)

Table 20 shows the MAPE resulting from using the developed regression models to estimate the total cost of 131 data sets of road construction projects. The results show the following:

- The MAPE of the developed models is ranging from 13% to 31% 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 use bid quantities as independent variables are more accurate than those using road length and road width as independent variables
- The table shows that for the models that use bid quantities as independent variables, when the number of work items involved in the model increase, the r^2 value increase and MAPE value decrease
- The table shows that for the models that use road length and width as independent variables, the MAPE is ranging from 18.8% to 31%. It shows that the model with highest accuracy is the model includes road length only with zero intercept value (model 8), while the model with least accuracy is the model includes road length and width (model 5)
- The table shows that for the models that use BOQ as independent variables, the MAPE is ranging from 13% to 19%. It shows that the model with highest accuracy is the model includes all construction activities (model 1), while the model with least accuracy is the model includes pavement quantity only (model 5).

Model No.	MAPE	r^2
1	13,3%	0,98
2	14,7%	0,97
3	16,1%	0,97
4	19%	0,95
5	31%	0,93
6	29,2%	0,95
7	27,1%	0,92
8	18,8%	0,95
9	23,4%	0,96
10	21,8%	0,92

Table 20 Mean absolute percentage error (MAPE) and r^2 of the developed regression models

Conclusion

This study aims at developing early cost estimating models for road construction projects using multiple regression techniques. The models were developed based on 131 set of data collected in the West Bank in Palestine. Such types of models are very useful, especially in its simplicity and ability to be handled by calculator or 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.

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

In this study, 10 regression models are developed; 4 of them include bid quantities as independent variable and 6 include road length and width. The coefficients of determination r^2 for the developed models range from 0.92 to 0.98. This indicates that the relationship between the independent and independent variables of the developed models is good and the predicted values from a forecast model fit with the real-life data.

The values of the mean absolute percentage error (MAPE) of the developed regression models are ranging from 13.3% to 31%. The results compare favorably with past researches which have shown that the estimate accuracy in the early stages of a project is between $\pm 25\%$ and $\pm 50\%$. The findings reveal that the models that use bid quantities as independent variables are more accurate than those use road length and road width as independent variables. But they require more information.

References

- Ahuja, H., Dozzi, S., and Abourisk, S. (1944) *Project Management Techniques in Planning and Controlling Construction Projects*. John Wiley and Sons, Inc., New York
- Bell, L. C. and Bozai, G. A. (1987) Preliminary cost estimating for highway construction projects, *AACE Transactions*, C.6.1–C.6.4
- Black, J. (1984) Application of parametric estimating to cost engineering, *AACE Transactions*, B.10.1–B.10.5.
- 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
- Gould, F. (2005) *Managing the construction process: estimating, scheduling, and project control*, Pearson Education, Inc., New Jersey
- Han, S. H., Kim, D. Y., and Kim, H., (2008) Two-staged Early Cost Estimation for Highway Construction Projects, Report (2008), Civil & 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
- Hendrickson, Chris and Au, Tung (1989) *Project Management for Construction: Fundamental Concepts for Owners, Engineers, Architects, and Builders*, 1st edition, Prentice Hall, London
- Holm, L., Schaufelberger, J., Griffin, D., and Cole, T. (2005) *Construction cost estimating: process and practices*, Pearson Education, Inc., New Jersey
- Humphreys, Kenneth K. (2005) *Project and Cost Engineer's Handbook*, 4th edition, Marcel Dekkar, USA
- Jelen, F. and Black, J. (1983) *Cost and Optimization engineering*, 2nd edition, McGraw-Hill, New York
- Lowe, D. J., Emsley, M. W., and Harding, A. (2006) 'Predicting construction cost using multiple regression techniques', *Journal of Construction Engineering and Management*, **132** (7), 750–758
- Mahamid, I. and Amund, A. (2010) 'Analysis of cost diverge in road construction projects', *Proceedings of the 2010 Annual Conference of the Canadian Society for Civil Engineering*, 9-12 June 2010, Winnipeg, Manitoba, Canada
- Mahamid, I. and Bruland, A. (2010) 'Preliminary cost estimating models for road construction activities', *Proceedings of the FIG Congress 2010 – Facing the Challenges – Building the Capacity*, 12-16 April 2010, Sydney, Australia
- Palestinian Economic Council for Development and Reconstruction (PECDAR), Personal communication, July, 2009, Ramallah, West Bank, Palestine.
- ROCKS (2002) Road Costs Knowledge System (v. 2.01), (The World Bank, Transport Unit – TUDTR). Retrieved August 4, 2010, from http://www.worldbank.org/transport/roads/rd_tools/rocks_main.html.

Rose, A. (1982) 'An organized approach to parametric estimating', Transactions of the seventh International Cost Engineering Congress. London, England

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

Schexnayder, C. J. and Mayo, Richard E. (2003) *Construction Management Fundamentals*, McGraw-Hill Higher Education, Boston, MA

Sodikov, Jamshid (2005) 'Cost estimation of highway projects in developing country: Artificial neural network approach', *Journal of the Eastern Asia Society for Transportation Studies*, **6**, 1036 – 1047.

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