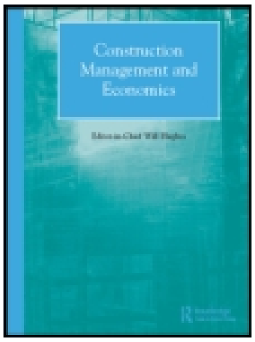

Construction Management & Economics

ISSN: 0144-6193 (Print) 1466-433X (Online) Journal homepage: https://www.tandfonline.com/loi/rcme20

Construction cost prediction for public school buildings in Jordan

Ayman H. Al-Momani

Construction Management and Economics (1996) **14**, 311–317

Construction cost prediction for public school buildings in Jordan

AYMAN H. AL-MOMANI

Civil Engineering Department, Mu'tah University, Karak, Mu'tah PO Box 7, Jordan

Received 2 May 1995; accepted 5 January 1996

An empirical model for predicting the construction costs for public school buildings is presented. Observations of 125 school projects were carried out in Jordan for the period 1984–1994. The results indicate that by the time a project is completed, the actual cost exceeds the original contract price by 30% while change orders result in an 8.3% cost overrun. Capital expenditure on school projects over the next 5 years is anticipated at 695.9 million JD. Out of this disbursement 208 million JD will be cost overrun and 57.7 million JD will result from issuing change orders. These figures highlight the need to improve current practices. A cooperative effort between concerned agencies is required to alleviate construction problems and could lead to new major innovations to meet the challenges ahead.

Keywords: Claims, cost, contract, planning, estimating, variations.

Introduction

The construction sector is one of the vital sectors in the development process of Jordan. The successful execution of construction projects, keeping them within estimated cost and the prescribed schedules primarily depends on the existence of an efficient construction sector capable of sustained growth and development in order to cope with the requirements of social and economic development and to utilize the latest technology in planning and execution.

The government contributes to the development of the construction industry in several ways. Many new construction projects are under way in Jordan. However, there are limitations and even drawbacks to these efforts. Both housing and public buildings experience delays in completion and constant modifications as work progresses. This has proven to be a serious and very expensive problem in Jordan's construction industry (Al-Momani, 1995). Alteration and modifications are consummated by formal change orders initiated by the owner or the engineer. These changes may alter the contract by additions, deletions or modifications to the work and will be added to the project cost and become an integral part of the final (actual) construction cost. Owners frequently find that the project cost greatly exceeds the budget and they are unable to identify the underlying reasons.

The literature search has revealed that some research work has been done in the past on the subject. Baldwin and Manthei (1971) examined the causes of delay in building projects in the United States. In their examination they showed that the major causes of delay are weather, labour supply and subcontractors. Chalabi and Camp (1984), on project delays in developing countries, indicated that adequate planning at the early stages of a project is crucial for minimizing delays and cost overruns. An empirical study carried out by Revay (1993) on 175 industrial, commercial, institutional and heavy engineering construction projects implemented during 1975–1991 in Canada, indicated $300 million of change orders exceeding the original contract price.

This trend is observable in Jordan and in many developing countries. In recognizing the magnitude of construction cost, it is important to determine the factors that are generating the cost overrun and to develop relationships that can be used in planning and estimating project cost. The data is part of a wider research study. Out of 125 projects costing over 28 049 993 JD (1 JD = $1.5), four projects were completed on schedule (see Figure 1). The rest are running behind schedule. These projects resulted in a total of 2 349 670 JD

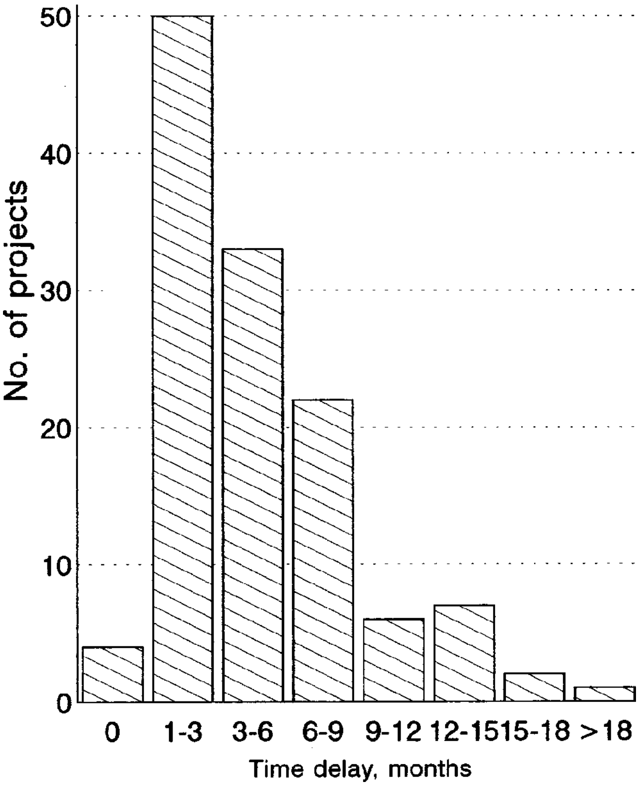

Figure 1 Length of time delays versus number of projects

worth of variation orders which represents 8.37% of the final project cost and is considered a factor of cost overrun. This can be explained by the number of change orders issued during construction and design revisions.

The objective of the present empirical study is, first, the identification and an understanding of the relative importance of major elements on the final cost of a construction project. Such a study is needed for both developing a prediction model for estimating the construction costs of building projects and as a basis for further investigation of cost in other types of projects. Second, with the important factors identified, the prediction model based on the observed data can be developed and used in planning. Third, problems can be identified before they occur, so that they can be avoided or their effects minimized. Fourth, major findings and recommendations are made for consideration by persons involved in construction management.

Data collection

The construction industry may be classified as heavy and building construction, where every project constitutes a heterogeneous group of activities. Because of the wide variety of construction projects, this study was limited to an examination of the construction costs of school buildings. Data for the predictor variables were obtained from Ministry of Public Works and Housing which includes public school buildings. The data provided detailed information on the cost of a project at awarding the contract, variation orders, the final area in square metres, the highest bidding cost, the actual completion date, the specified completion date and the length of time extensions. The number of projects selected for this study is 125 projects using annual observations from 1984 to 1994. This database appears to be representative for a large number of projects and for its coverage across the country.

Evaluation method

An understanding of construction processes and prediction of construction variables is becoming increasingly important for proper management. Construction management and process are a function of related variables such as planning, estimating, purchasing, quantity surveying and site management (Pasquire, 1991). However, the processes are even more complex. Fortune and Skitmore (1994) suggested that the project cost and price forecasts rely on the measurement and quantification of the construction work involved and are core skills of the construction manager. Further studies suggested that the large variances in costs or schedules will impact on the profitability, cash flow and, in extreme cases, the viability of projects (Al-Tabtabi and Diekmann, 1992). Estimating the cost of construction projects and then controlling the cost within a budget is an essential part of successful management. These factors have been repeatedly stated to be the outstanding need of construction in Jordan.

The construction process is subject to the influences of unpredictable factors, such as shortages in material, money, machines and manpower. Even if these processes were reasonably well understood up to a point, the heterogeneity and scale of the project make them extremely difficult to define theoretically. As a result construction managers often resort to empirical models relating construction cost variables to a project area (Hillebrandt, 1985; Ashworth, 1992; Bowen, 1982). One common expression is the regression equation with project area as the independent variable. In this evaluation, we needed to compare the final construction cost with several explanatory variables. This is the general viewpoint taken in this study. The theory is as follows:

$$Y = f(X_1, X_2, \ldots, X_n)$$

where Y equals the final construction cost and the Xs are all explanatory variables. In applying multiple linear regression (MLR), it is assumed the relationship between Y and the independent variables can be approximated by a linear model which provides best fit estimates of the model parameters by minimizing the error of the model (Draper and Smith, 1981). The multiple regression model for all assumed explanatory variables has the following form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon \qquad (1)$$

where Y is an observable random variable, $X_1, X_2, \ldots, X_p$ are known mathematical variables used to predict Y, the regression coefficients β_0, β_1, $\beta_2, \ldots, \beta_k$ are unknown parameters and ε is assumed to be a random, normally distributed error term. The least squares criterion can be used to determine the estimates for β_i, the error sum of squares,

$$\begin{aligned} SS_E &= \sum_i e_i^2 = \sum_i (Y_i - \hat{Y}_i)^2 \\ &= \sum_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \ldots - \hat{\beta}_p X_{pi})^2 \end{aligned} \qquad (2)$$

The error sum of squares must be minimized by setting $\partial\,(SS_E)/\partial\hat{\beta}_i = 0$, to obtain the system of the normal equation as follows:

$$n\hat{\beta}_0 + \hat{\beta}_1 \Sigma X_{1i} + \hat{\beta}_2 \Sigma X_{2i} + \ldots + \hat{\beta}_p \Sigma X_{pi} = \Sigma Y_i$$

$$\begin{aligned} &\hat{\beta}_0 \Sigma X_{1i} + \hat{\beta}_1 \Sigma X_{1i}^2 + \hat{\beta}_2 \Sigma X_{1i} X_{2i} + \ldots \\ &\quad + \hat{\beta}_p \Sigma X_{1i} X_{pi} = \Sigma X_{1i} Y_i \end{aligned}$$

$$\begin{aligned} &\hat{\beta}_0 \Sigma X_{2i} + \hat{\beta}_1 \Sigma X_{1i} X_{2i} + \hat{\beta}_2 \Sigma X_{2i}^2 + \ldots \\ &\quad + \hat{\beta}_p \Sigma X_{2i} X_{pi} = \Sigma X_{2i} Y_i \end{aligned}$$

$$\begin{matrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{matrix}$$

$$\begin{aligned} &\hat{\beta}_0 \Sigma X_{pi} + \hat{\beta}_1 \Sigma X_{1i} X_{pi} + \hat{\beta}_2 \Sigma X_{2i} X_{pi} + \ldots \\ &\quad + \hat{\beta}_p \Sigma X_{pi}^2 = \Sigma X_{pi} Y_i \end{aligned} \qquad (3)$$

To obtain the regression coefficients of $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p$, for Equation 1 we are required to solve the system of the $p + 1$ linear equation by the matrix notation where β_i are estimated by $S\hat{\beta} = X'Y$. Since computer programs are available, the author will not discuss the numerical solution nor the matrix notation. The sum of squares due to errors can be written as

$$\begin{aligned} SS_E &= \sum_i e_i^2 = \sum_i (Y_i - \hat{Y}_i)^2 \\ &= \sum_i (Y_i - \bar{Y})^2 - [\hat{\beta}_1 \sum_i (X_{1i} - \bar{X}_1)(Y_i - \bar{Y}) \\ &\quad + \ldots + \hat{\beta}_{pi} \Sigma (X_{pi} - \bar{X}_p)(Y_i - \bar{Y})] \end{aligned}$$

where $\bar{X}_j = \Sigma_{i=1}^{n} X_{ji}/n$. The sum in the brackets is the sum of squares due to regression, thus

$$SS_E = SS_{TC} - SS_R$$

$$SS_{TC} = SS_E + SS_R$$

where

$$\begin{aligned} SS_R = \hat{\beta}_1 \sum_i (X_{1i} - \bar{X}_1)(Y_i - \bar{Y}) + \ldots \\ + \hat{\beta}_{pi} \Sigma (X_{pi} - \bar{X}_p)(Y_i - \bar{Y}) \end{aligned} \qquad (4)$$

A hypothesis of interest is $H_0 : \beta_1 = \beta_2 = \beta_3 = \ldots = \beta_p = 0$. For testing this hypothesis, the table of analysis of variance is performed. The predictive performance of MLR models is judged by the coefficient of determination, R^2:

$$R^2 = \frac{SS_R}{SS_{TC}} (SS_{TC} - SS_E)/(SS_{TC} = 1 - \frac{SS_E}{SS_{TC}}$$

$$0 \leqslant R^2 \leqslant 1$$

The multiple linear regression appears to be appropriate for this type of problem. The specification of our model is determined by a number of independent variables which were thought to affect the dependent variable. The variables employed in the analysis are presented in Table 1. The estimation of the model allows us to analyse innovations and to show the structural interdependence of the project cost.

Analysis of the developed model

The relationship between the final construction cost and the explanatory variables was tested in several different ways in an effort to find the most definitive equation. Since the same data were used to compute several forms of linear equations, the procedure used for selecting the best form was to choose the form which gave the highest

Table 1 The variables employed in the analysis of construction cost

Y	=	Final construction cost
X_1	=	Cost of project at awarding the contract
X_2	=	Variation orders
X_3	=	Final area in square metres
X_4	=	Actual completion date
X_5	=	Specified completion date
X_6	=	Length of time extensions
X_7	=	Highest bidding cost
X_8	=	Lowest bidding cost

coefficient of determination, R^2 and the highest coefficient of multiple correlation, R, assuming that the errors are independently and normally distributed with mean zero and variance σ^2 and that the errors are homogeneous over the region of interest. If the distribution of errors is not normal or the variance is not homogeneous, it may be necessary to transform the variable to attain desirable properties. The first step in the analysis was to reduce the number of independent variables to a number that can represent the final construction cost adequately. Three of the explanatory variables were significant, while the insignificant variables were eliminated. The estimated linear equation for significant explanatory variables is

$$Y = 13\,300 + 0.732X_1 + 0.247X_2 + 42.8X_3 \qquad (5)$$

where Y is the total cost of the project, X_1 is the project cost at awarding a contract, X_2 is the variation order and X_3 is the area of the project. The multiple coefficient of determination $R^2 = 0.26$. We realized that a model needs an extremely high coefficient of determination (R^2) for it to predict the value of the dependent variables to high level of precision.

In order to use estimation techniques for the model parameters, logarithmic transformations of the variables are required to model the relationship between the variables adequately. In this study we found that the transformation produces a more linear relationship between these variables and a higher coefficient of determination. The function is of the following form:

$$Y_i = \beta'_o X_{1i}^{\beta 1} X_{2i}^{\beta 2} \ldots X_{ki}^{\beta k} \cdot \varepsilon_i \qquad (6)$$

where $\beta_o = \log \beta'_o$, Y_i is the final cost of the project and, ε is the random error of the transformed model.

After analysing the function form, it was decided that a log linear function best defined the relationship that existed between the dependent variable and the independent variables included in the model. This form is linear in logarithms and can be analysed with standard linear regression. After logarithmic transformation, the corresponding linearized form of the log model would appear as follows:

$$\log Y_i = \beta_o + \beta_1 \log X_{i1} + \beta_2 \log X_{i2} + \ldots \qquad (7)$$
$$+ \beta_k \log X_{ki} + \varepsilon$$

This method can be used to determine the parameters of β, thus providing minimum variance and unbiased estimates of these parameters. Using the linear model, Equation 7 results in an unbiased prediction of the dependent, Y, given the values of the Xs, since the expected value of the error term, ε, is zero and has no effect on the linear prediction equation.

Again, the data were pooled to arrive at a regression equation. The results of the equation relating to total investment costs for all projects with the three measures is given below:

$$\log Y_i = 1.574 + 0.277 \log X_{i1} + 0.03 \log X_{i2} \qquad (8)$$
$$+ 0.717 \log X_{i3}$$

or

$$Y_i = 37.49\, X_1^{0.277} X_2^{0.03} X_3^{0.717}$$

The R^2 for Equation 8 is 0.88 and all regression coefficients are significant at the 99.9% confidence level. The analysis of variance for the full regression is presented in the Appendix. After fitting the equation, we re-examined all the independent variables for any trends that might significantly enhance the performance of the model and found none. The p value of 0.00 indicates that we have a highly significant regression fit and it can be concluded that the original contract cost, variation order and project area contribute significantly to the regression model. For the entire sample, the estimated intercept has a positive sign and is very significant.

The simple correlation coefficients between the four variables and the associated test statistics are presented in the Appendix. Variables that are statistically independent will have an expected correlation of zero. As expected, the table does not indicate any zero correlations. The source of strong correlation may be explained by the fact that it is difficult to separate the effects of the total project cost from the effect of area.

Model verification

The developed model was constructed from 125 observations and will be used to predict the final cost using future data. This procedure is illustrated in Figure 2, using data for the actual and expected cost. It is not surprising that there is usually a close correlation between actual cost and expected cost, since the former is one component of the latter. It is a reasonable expectation that an increase in one will be accompanied by an increase in the other. Diagnostic statistics indicate neither the presence of heteroscedasticity nor autocorrelation. The small scatter of data around the theoretical line implies that the prediction process is in control and the model structure and parameters follow closely the process dynamics and stay the same during the prediction period. Therefore the model makes a satisfactory unbiased prediction.

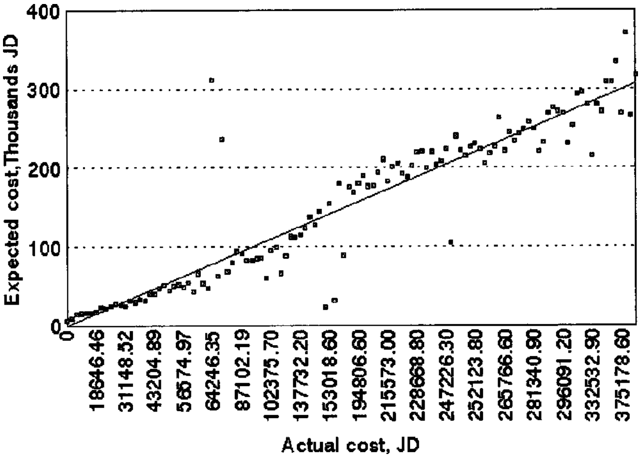

Figure 2 Actual cost versus expected cost

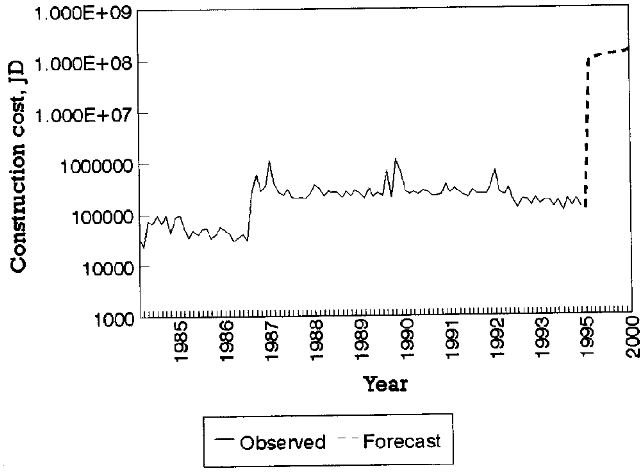

Figure 3 Projected construction cost

The preceding analysis has assumed that all projects are drawn from the same population, i.e. that all systems utilize similar projects (public schools) and technology. However, larger projects may differ substantially in the underlying technology from those projects that are under consideration. An assessment for the variables used in this model can be made by accepting the data presented in the Five Year National Plan For Economic and Social Development and the Annual Report of the Ministry of Public Works and Housing as planned government capital expenditure on school building during 1995–2000 which are shown in Table 2. A 5 year projection is shown in Figure 3 based on the data presented above. A 695.9 million JD of capital expenditure on school projects will occur in the next 5 years. Out of this disbursement 208 million JD will be as cost overrun and 57.7 million JD will result from issuing change orders. Such estimates give complete attention to current practices for improvement and help to determine the financial needs if the current practices continue.

Conclusion

This study presents a methodology that can be used in the analysis of final construction cost; while the methodology was applied to the problem of estimating school construction costs, the data analysis approach is also applicable for analysing residential and non-residential buildings. In applying the methodology to data from school projects, variation orders were taken into consideration as well as all reasonable combinations of variables which were thought to effect the total project cost.

Since the independent variables are the most useful and objective policy instruments available for construction managers, a reduction in variation orders reduces the need for additional capital expenditure. Complete drawings and bid documents, consistency in specifications and the use of a project cost system should become a necessity in improving productivity and can bring about a substantial effect on the construction sector. Therefore, the developed model provides a useful tool for decision makers wishing to investigate future alternatives and strategies and making rational decisions in determining the financial needs.

Achieving national objectives calls for a reduction of all construction problems. Efforts should be made to establish a financially viable industry with strong administrative and technically competent construction managers to identify areas of inefficiency and duplication in the construction sector and to propose improvement. There are many opportunities to minimize construction problems, but to take advantage of these opportunities adequate planning must be made and soon.

Table 2 Planned government capital expenditure on school buildings during 1995–2000

Year	Projected disbursement cost (JD)	Area (m^2)
1995	65 800 000	411 250
1996	71 000 000	443 750
1997	84 600 000	497 647
1998	92 650 000	514 722
1999	97 700 000	528 500
2000	116 700 000	583 500

References

Al-Momani, A. (1995) Construction practices: the gap between intent and performance, *Journal of Building Research and Information*, **23**(2), 87–91.

Al-Tabtabi, H. and Diekmann, J.E. (1992) Judgmental forecasting in construction projects, *Construction Management and Economics*, **10**, 19–30.

Ashworth, A. (1992) *Cost Studies of Buildings.* Longman, UK.

Baldwin, J.R. and Mathei, J.M. (1971) Causes of delay in the construction industry, *Journal of Construction Division, ASCE*, **97**(2), 177–87.

Bowen, P. (1982) Problems in econometric cost modeling, *The Quantity Surveyor.*

Chalabi, F.A. and Camp, D. (1984) Causes of delays and overruns of construction projects in developing countries, *CIB Proceedings*, **2**, 723–34.

Draper, N.R. and Smith, H. (1981). *Applied Regression Analysis* (2nd edn). John Wiley and Sons, New York.

Fortune, C. and Skitmore, M. (1994) Quantification skills in the construction industry, *Construction Management and Engineering*, **12**, 79–88.

Hillebrandt, P.M. (1985) *Economic Theory and the Construction Industry* (2nd edn) Macmillan Press Ltd, London.

Pasquire, C.L. (1991) Builder's quantities for contractors management. PhD thesis, Loughborough University, Loughborough.

Revay, S.G. (1993) Can construction claims be avoided?, *Building Research and Information*, **12**, 56–8.

Appendix

Summary of a regression analysis

Log transformation used on Y
The estimated regression coefficients are as follows:

$b_o = 1.574$
$b_1 = 0.277$ Log transformation used on X_1
$b_2 = 0.030$ Log transformation used on X_2
$b_3 = 0.717$ Log transformation used on X_3

Standard error of estimate 0.156
Standard error of (b_0) = 0.192 Computed T_o = 7.12451
Standard error of (b_1) = 0.069 Computed T_1 = 3.99081
Standard error of (b_2) = 0.029 Computed T_2 = 1.01282
Standard error of (b_3) = 0.076 Computed T_3 = 9.40052

Coefficient of determination $R^2 = 0.88512$
Coefficient of multiple correlation $R = 0.94081$
Coefficient of determination corrected for
degrees of freedom = 0.88227

Variable	Mean	Standard deviation	Correlation X versus Y
X_1	5.270	0.0390	0.879
X_2	3.900	0.053	0.580
X_3	3.124	0.037	0.929
Number of observations	125	p value: 0.000	F statistic: 310.7469

Source of variation	df	Sum of squares
Regression	3	22.943876
Residual	121	2.977995
Total	124	25.921871

Correlation matrix

	Y	X_1	X_2	X_3
Y	1			
X_1	0.879069	1		
X_2	0.580613	0.519589	1	
X_3	0.929828	0.882117	0.601286	1