A Cost Estimation Model for Highway Projects in Korea

Kim, Soo-Yong* Kim, Young-Mok** Luu, Truong-Van***

Abstract

Many highway projects are under way in Korea. However, owners frequently find that the project cost exceeds the budget and they are unable to identify the underlining reasons. The main purpose of this research is to develop cost models for transportation projects in Korea using the multiple linear regression (MLR). The data consist of 27 completed transportation projects, built from 1991 to 2001. The technique of multiple regression analysis is used to develop the parametric cost estimating model for total budget cost per highway square meter (TBC/m^2). Findings of the study indicated that MLR can be applied to highway projects in Korea. There are two major contributions of this research: (1) the identification of transportation parameters as a significant cost driver for transportation costs and (2) the successful development of the parametric cost estimating models for transportation projects in Korea.

Keywords: multiple regression; cost estimation; highway project; Korea.

1. Background and Objectives of study

"Cost estimating models are very useful at the early stages of a project's life cycle when little information is known about the project's scope" (Hegazy and Ayed, 1998). Therefore, it is important to develop a model to predict construction costs of infrastructure projects such as highway projects. Many highway projects are under way in Korea. However, owners frequently find that the project cost exceeds the budget and they are unable to identify the underlining reasons (Kim, 2008). The main purpose of this research is to develop conceptual cost model for total budget cost per square meter (TBC/m^2) of highway projects using the multiple linear regression (MLR). The prediction model may facilitate the government to evaluate and negotiate the construction cost of highway projects with contractors.

2. Methodology

2.1 Previous studies

Analysis of different classifications of estimates concludes that there are two main types of estimates: conceptual and detailed (Barrie and Paulson, 1992; Bley, 1990; Popham, 1996).

Traditionally, cost estimating relationships are developed by applying regression analysis to historical project information (Hegazy and Ayed, 1998). Regression analysis is a good method of determining the relationship between the parameters and cost, and determining the exact mathematical form of the model (Black, 1984; Orczyk, 1990).

Bley (1990) states that parametric estimating is one of the most used methods for conceptual cost estimating. Parametric estimating has been applied extensively across various industries including construction industry (Al Tabtabai et al., 1999; Hegazy and Ayed, 1998; Al Tabtabai et al., 1999; Wilmot and Cheng, 2003; Wilmot and Mei, 2005; Lowe et al., 2006).

2.2 Research methodology

This research adapted parametric cost estimating procedure proposed by Black (1984) as the research methodology. This procedure consists of six steps: (1) the definition of the problem encompassing the determination of the objectives and scope of the research study (Black, 1984); (2) data collection on design or engineering parameters that drive parametric cost estimates are developed from historical cost databases (Meyer et al., 1999); (3) the normalization of the data ensuring that every individual record in the database is

* 종신회원, 부경대학교 건설공학부교수
kims@pknu.ac.kr
** 일반회원, 코오롱건설 토목사업본부 부장
king2k@dreamwiz.com
*** 일반회원, 부경대학교대학원 건설관리공학협동과정 박사과정
luutruongvan@yahoo.com

on the same base; (4) the parametric cost estimating model development; (5) the establishment of model limitations; and (6) the documentation of the model development process.

The data consist of 27 completed transportation projects, commenced from 1991 to 2001. Seven projects were randomly selected for a validation data set. The technique of multiple regression analysis is used to develop the parametric cost estimating model for TBC/m^2 of highway projects

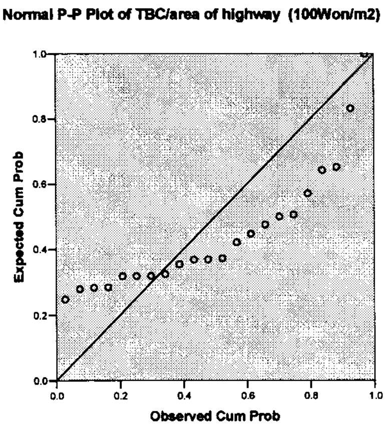
3. Model development

3.1 Variables Suggested in Previous Study

According to previous studies in highway projects, there are some parameters, which can be considered as independent variables (Creese and Li, 1995; Hegazy and Ayed, 1998; Al Tabtabai et al., 1999; Wilmot and Cheng, 2003; Wilmot and Mei, 2005; Lowe et al., 2006) such as length and width of highway and tunnel, the project purpose, total length of all piles, pavement thickness, bridge frame type, and design fee.

3.2 Assumptions for Multiple Regression Analysis

In order to fit a linear model, the underlying linear relationship is needed. To see whether the linearity assumption is reasonable, the Straight Enough Condition (SEC) is check for each independent variable. Scatterplots of dependent variables against each of predictors will indicate whether the SEC is satisfied. The results indicated that histograms of the dependent variable "TBC/m^2" is satisfied normality assumption. Normal, the normality of each variable is accessed through normal probability plots or Q-Q plots. The normality of TBC/m^2 is substantially improved after transforming (Figure 1 and Figure 2).



Fifure 1. Scatter plot of TBC/m^2 before transforming

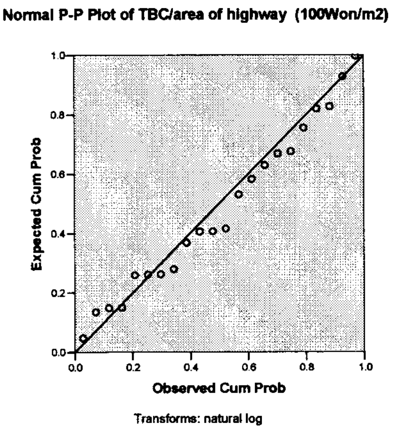


Figure 2. Scatter plot of TBC/m^2 after transforming

3.3 Multiple Regression Analysis

Multiple regression analysis is a statistical technique used to analyze the relationship between a single dependent variable and several independent variables. The basic formulation is:

$\hat{Y}$= b0 + b1*V1 +······+ bn*Vn

where: $\hat{Y}$=Predicted dependent variable; b0=Constant; b1=Partial regression coefficient of V1; V1=Independent variable 1; bn=Partial regression coefficient of Vn; Vn=Independent variable n; n=number of independent variables.

In order to determine whether or not the model with its independent variables is a significant predictor of the dependent variable (i.e., significant test for R^2), a statistical significance test is performed using the F-statistic. The R^2 value is statistically significant, or different from zero, if the computed F-statistic is greater than the F-critical value for the defined probability level. For this study, the p-value used is at a 0.05 level.

Stepwise regression is select to develop regression model. Stepwise regression includes regression models in which the choice of predictive variables is carried out by an automatic procedure.

In stepwise method, at each step, the independent variable not in the equation which has the smallest probability of F is entered, if that probability is sufficiently small. Variables already in the regression equation are removed if their probability of F becomes sufficiently large. The method terminates when no more variables are eligible for inclusion or removal. Therefore, significant variables are selected by the stepwise method

The stepwise regression resulted in the best model for $Ln(TBC/m^2)$ is (Table 1):

$$Ln(TBC/m^2)=8.508-0.02*\sqrt{Length\_of\_highway}+8.55*10^{-6}*Area\_tunnel$$

Table 1. Model summary for $Ln(TBC/m^2)$

Model Summary[c]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.736[a]	.542	.516	.40297721473	
2	.852[b]	.726	.694	.32069582594	1.556

a. Predictors: (Constant), SQRT(lengthP)
b. Predictors: (Constant), SQRT(lengthP), Area of tunnel (m2)
c. Dependent Variable: Ln(TBCnew/m2)

3.4 Multiple Regression Diagnostics

An important part of regression analysis is checking that the required assumptions are met. Residual analysis was performed to evaluate the assumptions of linearity, normality and homoscedasticity. The residuals are the differences between the observed and the predicted values. Homoscedasticity is a description of data for which the variance of the error terms appears constant over the range of values of an independent variable (Hair et al. 1998). Figure 3 and 4 confirmed that the required assumptions are met for the model developed. Moreover, a histogram of the standardized residuals and normal probability plot was run to check the linearity and normality assumption. Figure 5 shows standardized residual for $Ln(TBC/m^2)$ satisfying the linearity and normality assumption.

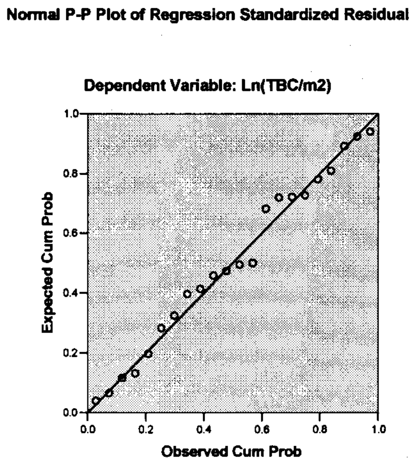


Figure 3. P-P plot of standardized residual for $Ln(TBC/m^2)$

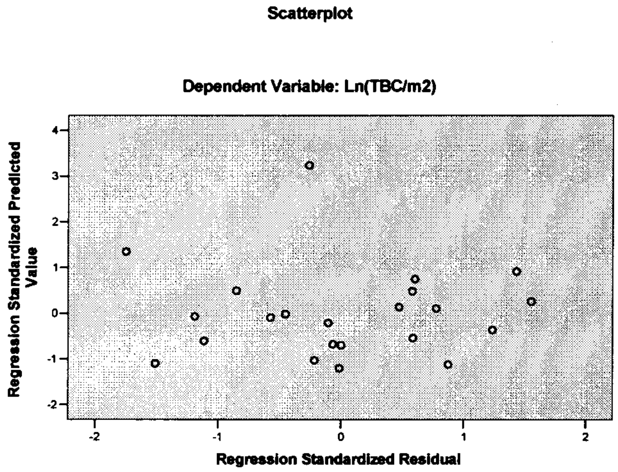


Figure 4. Standardized residual vs. standardized predicted value (TBC/m^2)

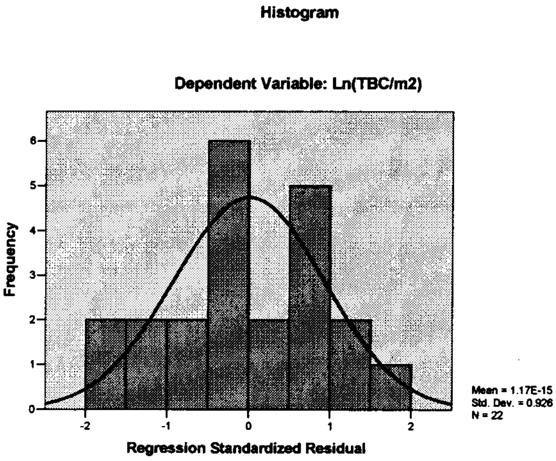


Figure 5. Histogram of standardized residual for $Ln(TBC/m^2)$

3.5 Model Validation.

Seven validation projects, representing about 24% of the population, are randomly selected and extracted from the data set. These validation projects are not used in any analysis steps, and are reserved for the subsequent model validation step.

The study adopted three relative measures of accuracy dealing with percentage errors pointed out in Goh (2000) to compare the forecasting results of the NN-based model. Those measures are percentage error (PE), mean percentage error (MPE), mean absolute percentage error (MAPE). The test resulted in MPE=-1.02% and MAPE=13.28%. This indicated that it is reasonably performed with the validation data set, with predictions ranging from underestimating by -30.24% to overestimating 22.30%.

4. Conclusions

This research identifies highway parameters that significantly influence the cost of highway projects. The data consist of 27 completed highway projects, commenced from 1991 to 2001. Seven projects were randomly selected for a validation data set. Twenty projects were used as the data set for regression. MLR is used to develop the parametric cost estimating model by establishing the cost estimating relationships between the highway parameters and TBC/m^2.

The model of TBC/m^2 pointed out that "Length of highway" and "Area of tunnel" are factors have strong effect to conceptual costs of highway projects in Korea, and indicated that MLR can be applied to highway projects in Korea. The absence of other variables does not imply that they have no effect to TBC/m2. A possible explanation is that the relation between other factors and TBC/m^2 may be non-linearity.

For further studies, there are needs to identify other parameters of highway projects and to perform additional data collection.

References

1. Al Tabtabai, H., Alex, A. P., and Tantash, M. (1999). "Preliminary Cost Estimation of Highway Construction Using Neural Networks", Cost Engineering, Vol. 41, No. 3, p. 19-24.
2. Black, James H., (1984). "Application of Parametric Estimating to Cost Engineering", AACE Transactions, Paper B.10.
3. Creese, R., and Li, L., (1995). "Cost Estimation of Timber Bridges Using Neural Networks", Cost Engineering, Vol. 37, No. 5, p. 17-22.
4. Goh, B.H (2000). "Evaluating the performance of combining neural networks and genetic algorithms to forecast construction demand: the case of Singapore residential sector." Construction Management and Economics, 18(2), pp.209-217.
5. Hair, Joseph F., Anderson, Rolph E., Tatham, Ronald L., and Black, William C. (1998). Multivariate Data Analysis, 5th Edition, Prentice-Hall, NJ.
6. 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.
7. Kim, Y.M. (2008). "Construction cost estimating for highway by multiple regression in Korea." Dissertation presented to the Pukyong National University in partial fulfillment of the requirements for the degree of Doctor of Philosophy, Busan, Korea.
8. Lowe, D.J., Emsley, M.W., and Harding, A. (2006). "Predicting construction cost using multiple regression techniques", Journal of Construction Engineering and Management, ASCE, Vol. 132, No. 7, p.p. 750-758.
9. Meyer Edward R., and Burns, Thomas J. (1999). "Facility Parametric Cost Estimating", 1999 AACE International Transactions, Paper EST.02.
10. Orczyk, Joseph, and Chang, Luh-Maan (1990). "Parametric Project Scheduling of Milestones", AACE Transactions, 1990.
11. Popham, Kymberli (1996). "Cost Estimating Using Historical Costs", 1996 AACE Transactions, Paper EST.14.
12. Wilmot, C.G., and Cheng, G. (2003). "Estimating future highway construction cost", Journal of Construction Engineering and Management, ASCE, Vol. 129, No. 3, 272-279.
13. Wilmot, C.G., Mei, B. (2005). "Neural network modeling of highway construction costs." Journal of Construction Engineering and Management, 131 (7), 765-771.