



**Institute of Internet and Intelligent Technologies**  
 Vilnius Gediminas Technical University  
 Saulėtekio al. 11, 10223 Vilnius, Lithuania  
<http://www.isarc2008.vgtu.lt/>

**The 25<sup>th</sup> International Symposium  
 on Automation and Robotics in Construction**

**June 26–29, 2008**

**ISARC-2008**

## **TWO-STAGED EARLY COST ESTIMATION FOR HIGHWAY CONSTRUCTION PROJECTS**

**Du Yon Kim**

Ph. D. Candidate, Department of  
 Civil & Environmental Engineer-  
 ing, Yonsei University  
 Seoul, Korea  
[cagedbird@yonsei.ac.kr](mailto:cagedbird@yonsei.ac.kr)

**Byoungil Kim**

Master Student, Department of  
 Civil & Environmental Engineer-  
 ing, Yonsei University  
 Seoul, Korea  
[ikim@yonsei.ac.kr](mailto:ikim@yonsei.ac.kr)

**Seung Heon Han**

Associate Professor, Department  
 of Civil & Environmental Engi-  
 neering, Yonsei University  
 Seoul, Korea  
[shh6018@yonsei.ac.kr](mailto:shh6018@yonsei.ac.kr)

### **ABSTRACT**

The quality of early cost estimates is critical to the feasibility analysis and budget allocation decisions for public capital projects. Various research have been attempted to develop cost prediction models in the early stage of a construction life-cycle. However, existing studies are limited on its applicability to actual projects because they focus primarily on a specific phase as well as utilize restricted information while the amount of information collectable differs from one another along with the project stages. This research aims to develop two-staged cost estimation model for the schematic planning and preliminary design process of a construction projects, considering the available information of each phase. In the schematic planning stage where outlined information of a project is only available, the case-based reasoning model is used for easy and rapid elicitation of a project cost based on the extensive database of more than 90 actual highway construction projects. Then, the representing quantity-based model is proposed for the preliminary design stage where more information on the quantities and unit costs are collectable based on the alternative routes and cross-sections of a highway project. Real case studies are used to demonstrate and validate the benefits of the proposed approach. Through the two-stage cost estimation system, users are able to hold a timely prospect to presume the final cost within the budge such that feasibility study as well as budget allocation decisions are made on effectively and competitively.

### **KEYWORDS**

Early Cost Estimate, Highway Construction Project, Case-based Reasoning, Regression Tree, Quantity-based model

### **1. INTRODUCTION**

Early cost estimates are critical factors to the initial decision making process for capital projects. From

the perspective of the infrastructure providers, accurate cost estimations allow for the effective budgeting and financing of a project. As such, the importance of early estimates to owners cannot be

overemphasized. Early estimates are typically plagued by limited scope definition (and thus high potential for scope changes) and are often prepared under stiff time constraints. Furthermore, reliable cost data are frequently under attack to obtain during the conceptual stages of a project, particularly where basic route design and geographic site conditions of a road remain unresolved. Early estimates, even when grossly inaccurate, often become the basis upon which all future estimates are judged.

This research aims to develop two-staged cost estimation model for the schematic planning and early design process of a highway projects, considering the available information of each phase. In the schematic planning stage where outlined information of a project is only available, the case-based reasoning model is used for easy and rapid elicitation of a project cost based on the extensive database of more than 90 actual highway construction projects. Then, the representing quantity-based model is proposed for the preliminary design stage of a project where more information on the quantities and unit costs are collectable based on the alternative routes and assumed cross-sections of a highway project.

## 2. LITERATURE REVIEW

The conceptual cost estimates for capital projects has been a major concerned and a subject for both the project owners and contractors. Hackney [3] published a checklist for establishing a detailed definition rating for capital projects. The study proposed the use of the definition checklist for applying contingency to capital cost estimates and then validated the checklist by comparing the definition ratings of 30 projects to their respective levels of cost overrun. Hegazy and Ayed [5] developed a neural network model for parametric cost estimation of highway projects. They considered characteristics of early cost estimates that conceptual estimation process is determined by some parametric values. By using this concept, they derived parametric values related with cost such as construction region, total length of highway.

Kim et al. [6] applied hybrid models of neural networks (NN) and genetic algorithms (GA) to cost estimation of residential building to predict preliminary cost estimates, based on the data samples con-

structed from 1997 to 2000 in Seoul, Korea. Lowe et al. [8] 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 such as gross internal floor area (GIFA), function, duration, mechanical installations, and piling.

Oberlender and Trost [9] concentrated more on improving prediction accuracy by deriving four determinants influencing the accuracy of an early cost estimates. Following this, they established a multiple regression model which derived significant relationship between estimate accuracy and influencing factors such as basic process design, team experience and cost information, time allowed to prepare the estimate, site requirements, and bidding and labor climate [11].

While the various methods have been used for cost predictions in its early stage, existing studies are limited on its applicability to actual highway projects, because they focus more on a specific phase as well as utilize restricted information while the amount of information collectable differs from one another along with planning and preliminary design stage. To this end, we investigate the actual budgeting process in the highway construction project, under the research collaboration of the Korean Ministry of Construction and Transportation. Then, we develop 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 budgeting and initiation.

## 3. FRAMEWORK FOR TWO-STAGED COST ESTIMATION

In Korea, the 'Highway Operation Handbook,' published by the Korean Ministry of Construction and Transportation is widely used as the standard to estimate the initial project costs of a highway infrastructure. The 'Highway Operation Handbook' defines the design codes and specifications of national express highway (1st grade of highway, toll roads) and the national general highway (2nd grade of highway, free roads). It also provides the standard

cost per unit length based on simple statistics from the completed road projects during a last three years (between 2003 and 2005). In addition it further classifies the cost of unit length in association with the general part of road embankment, bridge, and tunnel sections, respectively. All together, it provides the unit costs under the categories of either new construction or expansion with each having the different lanes. Table 1 shows the estimation standard of highway construction costs in this Handbook.

Table 1. Estimation Standard of Construction Costs [7]

Section		General parts (embankment, bridge, tun- nel)	Bridge	Tunnel
Express highway	4-lanes extension	11.3million USD/km	49.2 mil- lion USD/km	26.7 mil- lion USD/km
	New 4- lanes	13.6 million USD/km	45.1 mil- lion USD/km	26.7 mil- lion USD/km
National highway	2-lanes extension	9.5 million USD/km	42.8 mil- lion USD/km	24.2 mil- lion USD/km
	New 4- lanes	10.0 million USD/km	44.5 mil- lion USD/km	25.1 mil- lion USD/km

As shown in Table 1, the current standard calculates the construction costs by the unit cost of every kilometer, primarily assuming the total estimation is simply in proportion to the road lengths regardless of its unique characteristics along the road routes. In this regards, it does not consider diverse variables of the construction traits such as a different geographic location, a construction period, percents of bridge or tunnel sections, and available important information, which differs from the every step of the project progress.

Meanwhile, the current process of a road design consists of a schematic planning and a design as depicted the Table 2. The schematic planning conducts the budget allocation through a preliminary feasibility study. The design phase has three more steps; a selection of the best route among the alternatives, a detailed field survey and underground tests, and a development of drawings & quantity take-offs.

The selection of a best route determines the optimal road route through the comparisons of competing alternatives. If the route is selected, the detailed field survey and underground investigations confirms conditions of the soil and ground by which a path, structure types and other design factors can be decided. Upon the field survey, the drawing work and the estimation process based on the quantity take-offs calculate the ultimate construction costs.

Table 2. Current Process of the Road Construction Project

Phase		Objective	Current Cost Est. Method
Schematic Planning		<ul style="list-style-type: none"> <li>Preliminary Feasibility study</li> <li>Project selection</li> <li>Alternative routes plan</li> </ul>	Unit cost
Design	Evaluating & Selecting a best route	<ul style="list-style-type: none"> <li>Determination of the best route</li> <li>Structure location/type</li> </ul>	Unit cost
	Detailed Field survey	<ul style="list-style-type: none"> <li>Earth work plan</li> <li>Path/pavement design</li> <li>Basic design of structures</li> </ul>	Unit cost
	Drawing & quantity estimation	<ul style="list-style-type: none"> <li>Details of design</li> <li>Estimation of final construction cost</li> </ul>	Quantity-based

Considering the user's perspectives, the effectiveness of the cost estimation model goes down after performing the detailed field surveys and drawing works, because it simply relies on the quantities and each unit cost to estimate the final cost when we have enough information on the cost and quantity from the drawings. Thus, it appears that the cost prediction models focusing on the 'schematic planning' and 'initial stage of a design during the routes selection', which are currently calculated only by the unit cost of every kilo-meter, are to maximize its usage.

As each step goes up, a decision maker has different information level. The schematic planning level entails rather basic information such as a size of the target project and other overviews. Then the further steps like the route selection and the field surveys

develop more detailed information. As Han et al. [4] mentioned, a tool development to support a prompt decision making within limited information meets

decision maker's demand. To this end, this research suggests the two-staged cost estimation process as depicted in Figure 1.

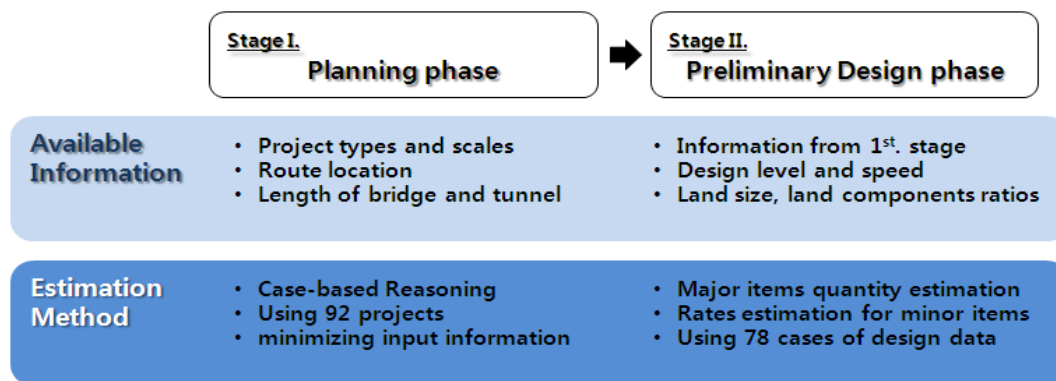


Figure 1. Frame work for two-staged cost estimation process

Stage I, schematic planning phase uses information of an initial level such as project type & size, lane location, This stage applies the case-based Reasoning (CBR), which estimates construction costs based on the previous similar projects. Stage II is modeled to select a best route among the alternatives. It derives the determining/representative items and their assumed quantities based on the various modeling techniques by using more detail information on the routes.

#### 4. STAGE I: PLANNING PHASE

Cost estimation model for the schematic planning phase applies the case-based Reasoning (CBR) method. CBR is originated from cognitive science by Schank and Abelson in 1977 and plays an important role as a decision supporting tool with artificial intelligent method [10]. Its process is similar to human decision development which is often based on the results from useful previous cases. In other word, when someone decides a solution to the problem, they rely on their own previous cases which are similar to the present case. As they compare the present case with a load of previous cases in many criterions, they gradually could search the most suitable case. To construct the model on the stage 1, this study collected 92 road project cases, delivered by

the government agency (Public Procurement Service) between 2005 and 2006. Throughput the data mining and fitting, the following eight cost influencing factors are drawn considering the available information level for this stage; a location, a project type (new, extension), a contract type (lowest price, pre-qualified competition, design-build), a construction period, total length of road, a road width, total length of bridge, and total length of tunnel.

These eight factors are standards to examine the similarity of each case and to estimate a construction cost by the CBR algorithm. In order to assess the similarity score for 92 cases, a weight calculation for the eight factors is required. This study applied a series of iterations based on the linear programming and subsequently, drew the optimum weights to minimize errors between expected value and real value of the costs of each sample (Table 3).

Based on the results as Table 3, this study validated the proposed CBR model using additional five road projects out of the sample data. As a result, the error rate was estimated at 18.4% in average which is highly accurate than 38.9% from the current method by the government handbook. In addition, it was found that the proposed CBR model brought more stable results for various cases contrary to the cur-

rent approach, showing a standard deviation only 5.2% error rate.

Table 3. Weight estimation results of the cost influencing factors

Cost influencing factor	Weight
Location	0.026
Project type	0.204
Contract type	0.098
Construction period (day)	0.156
Total length of road (kilo-meter)	0.025
Road width (meter)	0.139
Total length of bridge (meter)	0.114
Total length of tunnel (meter)	0.128

## 5. STAGE II: PRELIMINARY DESIGN PHASE

Stage II requires more accuracy of cost estimation in order to compare the feasibilities of each route alternative. Therefore, by using more detail information collected, this paper applies a representing/determining quantity-based estimation model through which the major item's estimated quantities are multiplied by each unit cost. Since it is not appropriate to predict every item's quantities, this research selects a representative item from each work type based on the correlation analyses of each item on the total cost. Then, the residual minor items are accounted by the pro-rates that are modeled based on the previous data sample. In this stage, we collected 78 cases of highway construction projects which were designed since 2000. Total 70 cases are used to develop models, and the rest of 8 cases are used to validate the models.

Generally, a highway construction work consists of the earthwork, slope stabilization works, drainage works, structures, tunnels, pavements, safety facilities, and other appurtenant works. Through the cost correlation analysis of the collected data, we selected earth work, drainage work, and pavement as the most prevailing work types to build the cost estimation models. Other works are estimated by lump-sum pro-rates. The research performs the structure works and the tunnel constructions that require additional analyses according to the various bridges and tunnel types. However, the structure and

tunnel works are excluded in this paper for the sake of brevity.

As results, we derived the total of 44 representing items that account for only 10% of the total number of items while covering the 69.6% of total cost. Each representing item was modeled by the appropriate technique to presume the respective quantity without any further information from the drawing. For example, it is a common practice that the earth work and the side ditch work do not have the standard drawings, but their quantities are influenced by several specified variables. To identify those variables and develop the estimation models for these items, this study applied CART (Classification and Regression Tree). CART is a way to classify items from high and low homogeneities [2]. For the earth work, it developed a quantity estimation model of the cutting and the filling of earth as the Table 4. The carrying of earth and the banking apply an arithmetic relationship between the cutting and the filling. In the case of the side ditch work, it is highly co-related to the quantity of the earth work. Therefore to develop the quantity estimation models of the drainage work, the quantities of the earth work are used as independent variables.

Table 4. Quantity estimation through work types

Work type		Quantity estimation method
Earth work		<ul style="list-style-type: none"> <li>Quantity estimation models are developed</li> <li>Using CART method</li> <li>Independent variable: project characteristics</li> <li>Dependent variable: cutting and filling of earth</li> </ul>
Drainage work	Side ditch	<ul style="list-style-type: none"> <li>Quantity estimation models are developed</li> <li>Using CART method</li> <li>Independent variable: project characteristics and quantity of earth work</li> <li>Dependent variable : length of side ditch</li> </ul>
	Culvert	<ul style="list-style-type: none"> <li>Unit length quantities are derived</li> <li>Using standard section</li> </ul>
Pavement		<ul style="list-style-type: none"> <li>Unit length quantities are derived</li> <li>Using standard section</li> </ul>
Slope stabilization work		<ul style="list-style-type: none"> <li>Applying respective rates of cost</li> </ul>
Safety facility work		<ul style="list-style-type: none"> <li>Applying respective rates of cost</li> </ul>
Appurtenant work		<ul style="list-style-type: none"> <li>Applying respective rates of cost</li> </ul>

Meanwhile, the culvert work and the pavement have linear characteristic. It is able to estimate the quantity of unit length from the standard section for the culvert work and the pavement. Therefore, it developed a model to estimate quantity through multiplying unit length quantity by total length of each work type. As portions of the slope stabilization work, the safety facility work, and the appurtenant work are relatively small in the whole project costs, estimations are applied respective rates. The rates of each work type are considered the characteristics of the projects through co-relation analysis.

In order to validate second level model, based on the major items' estimations, randomly picked 8 among the 78 cases are subjects to compare an expected and real design values. As the Table 5 shows, the earth work indicates 7.85% of an average error rate, and the drainage work, and pavement have -1.78%, -7.12%, respectively. Generally the result has an excellent precision considering that the accuracy of predicting costs in the preliminary design level that the AACE (American Association of Cost Engineers) proposes is -20%~+20% [1].

Table 5. The validation result of 2<sup>nd</sup> phase model

Work type	Average error rate	Standard deviation error rate
Earth Work	7.85%	46.74%
Drainage Work	-1.78%	41.91%
Pavement	-7.12%	37.84%

## 6. LESSONS LEARNED & CONCLUSION

The study proposed the phased cost estimation process of road construction projects for a practical application in order to make a reasonable estimation of public road construction. The derived model consists of two stages considering available information at the corresponding moment. Compared the two-staged estimation process to the previous methods and the present standard, it shows an excellent level of accuracy. It means the result could be applicable in the future. If the additional methods such as estimations of bridges and tunnels complement the study model, it would propose the more in-depth system for road construction estimations.

## ACKNOWLEDGEMENT

This research was supported by a grant (06 CIT A03) from Construction Infra Technology Program funded by Ministry of Land, Transport and Maritime Affairs.

## REFERENCES

- [1] AACE (1997), Recommended Practice No.17R-97 (1997): Cost Estimate Classification System, AACE, Inc.
- [2] Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984) Classification and Regression Trees, Chapman & Hall (Wadsworth, Inc.): New York.
- [3] Hackney, J. W. (1985) Applied contingency analysis, Transactions of the American Association of Cost Engineers, Vol. B, No. 2, pp. 1-4.
- [4] Han, S. H., Kim, D. Y., and Kim, H. (2007) Predicting Profit Performance for Selecting Candidate International Construction Projects, Journal of Construction Engineering and Management, Vol. 133, No. 6, pp. 425-436.
- [5] Hegazy, T., and Ayed, A. (1998) Neural Network Model for Parametric Cost Estimation of Highway Projects, Journal of Construction Engineering and Management, Vol. 124, No. 3, pp. 210-218.
- [6] Kim, G. H., Seo, D. S., and Kang, K. I. (2005) Hybrid Models of Neural Networks and Genetic Algorithms for Predicting Preliminary Cost Estimates, Journal of Computing in Civil Engineering, Vol. 19, No. 2, pp. 208-211.
- [7] KMOCT (2006) Road Operation Handbook 2006, Korea Ministry of Land, Transport and Maritime Affairs.
- [8] Lowe, D. J., Emsley, M. W., and Harding, A. (2006) Predicting Construction Cost Using Multiple Regression Techniques, Journal of Construction Engineering and Management, Vol. 132, No. 7, pp. 750-758.
- [9] Oberlender, G. D., and Trost, M. (2001) Predicting Accuracy Early Cost Estimates Based on Estimate Quality, Journal of Construction Engineering and Management, Vol. 127, No. 3, pp. 173-182.
- [10] Riesbeck, C. K., and Schank, R. C. (1989) Inside Case-based Reasoning, Lawrence Erlbaum.
- [11] Trost, S. M., and Oberlender, G. D. (2003) Predicting Accuracy of Early Cost Estimates Using Factor Analysis and Multivariate Regression, Journal of Construction Engineering and Management, Vol. 129, No. 2, pp. 198-204.