SCHEMATIC ESTIMATING MODEL FOR CONSTRUCTION PROJECTS –USING PRICIPLE COMPONENT ANALYSIS AND STRUCTURAL EQUATION METHOD

Young-Sil Jo[1], Hyun-Soo Lee[2] and Moon-Seo Park[3]

[1] Master, Construction Engineering & Management, Seoul National University Korea
[2] Professor, Construction Engineering & Management, Seoul National University, Korea
[3] Professors, Construction Engineering & Management, Seoul National University Korea
Correspond to ysjo02@nate.com

ABSTRACT: In the construction industry, Case-Based Reasoning (CBR) is considered to be the most suitable approach and determining the attribute weights is an important CBR problem. In this paper, a method is proposed for determining attribute weights that are calculated with attribute relation. The basic items of consideration were qualitative and quantitative influence factors. These quantitative factors were related to the qualitative factors to develop a Cost Drivers-structural equation model which can be used to estimate construction cost by considering attribute weight. The process of determining the attribute weight-structural equation model consists o 4 phases: selecting the predominant Cost Drivers for the SEM, applying the Cost Driers in the SEM, determining and verifying the attribute weights and deriving the Cost Estimation Equation. This study develops a cost estimating technique that complements the CBR method with a Cost Drivers-structural equation model which can be actively used during the schematic estimating phases of construction.

Keywords: Construction costs, Schematic Estimating, Case-based reasoning (CBR), Influence factors, Attribute weight, Structural Equating Model (SEM)

1. INTRODUCTION

The construction industry is a 'job order production' industry because it makes products that are distinctive and particular compared to the products' of other industries. In other words, not only is 'uniqueness' the most fundamental characteristic of the construction industry, but it is also a project's 'uniqueness' that directly impacts the establishment of any sort of regulated construction cost range. Therefore, future planning for any construction project is based on knowledge and information gathered from past experiences (Yau, 1998). Although there are various methodologies that utilize past construction project data, for considering the 'uniqueness' factor inherent in the construction industry, Case-Based Reasoning (CBR) is considered to be the most suitable approach (David, 2006). CBR is a reasoning method that modifies past experiences accordingly to find answers to a current problem. The CBR estimation system's efficiency is impacted by 6 elements: (1) the composition method of the case-base, (2) the attributes use in estimation, (3) the attribute weights, (4) the function used in similarity estimation, (5) the number of nearest neighbors, and (6) the generation method of estimation. Thus, to address the construction industry's distinctive characteristics, it is crucial to build a successful CBR system that can determine the most similar case. Cost influence factors are used to search for similar case standards. The core of this process (i.e., searching for similar cases) is the retrieve, reuse stage. Weighting the relative important of these attributes is moiré difficult. In fact, determining the weights if the attributes is an important CBR problem (Sevgi, 2008).

As demonstrated by Brown and Gupta (1994), CBR has been actively applied to construction cost estimation and to the area of performance measurement. As well, there are currently several methods to determine attribute weight such as Ann, AHP, decision tree, and logistic regression analysis, all of which could generally optimize the CBR system. To solve the problems of selecting attributes and determining attribute weight, Aha and Bankert (1994) and Jon (2001) have removed the low correlation attribute and have applied the proper weight that is determined by valuation of the estimation attribute. This research has improved the model of prediction performance. Furthermore, Park and Han (2002) have determined weight using AHP and have reflected weights in the CBR system as input data in an attempt to improve its prediction performance. In order to improve the prediction performance of CBR, Ahn and Kim (2006) have used ANN to develop optimization techniques in which weight and selection case are performed simultaneously. On the other hand, Sevgi and David (2008) have improved cost estimation accuracy by utilizing the decision tree while determining weight. Furthermore, a Logistic Regression Model makes modeling easy, and it can handle the calculation of the statistical significance. However, with this type of model, it is difficult to analyze the complicated non-linear correlation assisted in the factors. Compared with statistical methods, the ANN model has different merits such as short analysis time and expense reduction. But, the ANN model is also difficult to verify because its results are determined through the attribute weights in a network area. On the other hand, Decision tree is strong in terms of prediction and verification because its prediction is performed through a matrix. However, the non-linear and non-safety characteristics of this particular method could potentially lead to misclassification when analyzing the linear correlations. Finally, not only can AHP determine weight and priority by using pair wise comparison, but it can also immediately assess quantitative and qualitative influence factors, thus ensuring that experts can access this data easily. However, the result standard of estimate in AHP is slightly different from regression analysis.

In this research, a method is proposed for determining attribute weights that are calculated with attribute relation. Such a method could have an absolute impact on the CBR system by considering various weights and based on apartment housing construction, which comprises a large portion of the domestic construction industry

2. SCHEMATIC ESTIMATING AND METHOD

2.1 Schematic Estimating

In general, a construction project consists of planning, designing, procurement, execution, completion, and transfer (in this sequence). As well, the design phase can be further divided into the schematic design and detail design phases. While cost estimation is conducted leading up to the construction phase, from the construction phase onward, cost estimation is considered to be cost management. A further elaboration on the separate construction project phases is as follows: (1) during the ‘feasibility and conceptual design phase’, construction cost can be assumed based on the data of similar previous construction projects. (2)The schematic design phase is also a phase that can refer to past experience, albeit limitedly, but this phase, depending on the method of execution, may provide timing for pivotal decision-making. (3) Because more detailed cost estimates can be determined during the actual construction phases, it can be argued that the connection between decision- making and reference to previous projects is relatively weak. Therefore, this paper defines 'schematic estimating' as the process of estimating construction cost which is only conducted previous projects is relatively weak. Therefore, this previous projects is relatively weak. Therefore, this paper defines 'schematic estimating' as the process of estimating construction cost which is only conducted during the ‘feasibility and conceptual design phase’ and the schematic design phase.

2.2 Schematic Estimating Methods

The schematic estimating methods proposed in domestic and international research can be primarily divided into four different methods: (1) a method that uses cost per pyeong (a unit of area) as the standard; (2) a method that uses a statistical process; (3) a method utilizing neural networks to determine the connection between construction costs and artificial intelligence; and (4) a method that analyzes the change in quantity to estimate costs. Table 1 discusses the pros and cons of these 4 methods.

In the statistical methods, a quantitative analysis of influence factors must be conducted, which, in turn, requires a considerable amount of statistical data. Also, in the case of a gradual increase in the diversification of influence factors, actively responding to each factor and estimating the cost of construction will become more difficult (Sevgi, 2006). On the other hand, the methods that rely on artificial intelligence allow for a more precise estimation than the statistical approach, even when the number of influence factors is low, because this particular method reduces the room for error through learning from examples (Ku, 1999). This method can be further divided into methods that use neural networks (ANN), genetic algorithms (GA), and Case-Based Reasoning (CBR).

On one hand, the method using neural networks (ANN)

Table 1. Methods of Cost Estimating

Class	Standardized method	Statistical method	AI method	Quantity change analysis
Type	□Physical dimensions method (AACE)	□ Method using recurrence (McCaffer 1975) □Monte carlo simulation (Baek1997)	□ ANN (Adeli 1998) □ GA (Miller 1989) □ CBR (Watson1995)	□ Quantity Based Active Schematic Estimating(Q-BASE, 2005)
Pros	□Prompt, convenient	□ Can decrease room for error because it is based on a mathematical model	□ More accurate than statistical method □ CBR easy to create & maintain precise models	□ Can evaluate credibility □ Quick response to change
Cons	□ Possibility of error □ Cannot be applied to complicated models	□ Inept to change due to time □ Problem of credibility within lineal relations	□ ANN is a ‘black box’ □ Deciding on ANN parameters is time consuming	□ Requires much time and effort □ Quantity estimates rely on recurrence only from similar previous projects

exhibits a 'black box' characteristic, which means that even though the method’s assessments of construction costs are relatively accurate, it is difficult to develop this method into information that is usable for future purposes? On the other hand, although the case-based reasoning method requires many various past examples and experiences, it is very useful for research purposes when there is an abundance of examples such as apartment housing projects (Sohn, 2005).

Thus, when taking the scope of this paper and the unique characteristics of the construction industry into account, the case-based reasoning (CBR) method is the most suitable approach for effective and accurate cost estimation. Recently, the CBR method has been utilized in various studies conducted in the field of construction, including the areas of architectural design, structural design, and construction management.

3. METHODOLOGY

3.1 Structural Equation Model(SEM)

A structural model is a tool that defines the connection between variables and analyzes these connections with a series of equations. This method is currently being applied actively in the fields of science and social science. Structural equations consist of latent variables, observed variables, exogenous variables, endogenous variables, and error variables. (1)Latent Variable: A variable that cannot be directly observed or measured. Latent variables cannot be measured in themselves and therefore are usually measured indirectly through observed variables. (2)Observed Variable: A variable that can be measured directly. Observed variables are sometimes measured by being linked with latent variables. (3)Exogenous Variable: As an independent variable, it influences different variables. (4)Endogenous Variable: A variable that is directly or indirectly influenced by other variables at least once within a structural equation model. (5)Error Variable (Measurement error): This shows the extent of the inability to completely explain the latent variables. (6)Structural error: A structural error occurs when a endogenous variable cannot be explained by one or more exogenous variable(s).

Figure 1 SEM using Construction factors

The structural equation model has several merits. First, it can be applied to regression analysis, primary factor analysis, relation analysis, and other analyses all simultaneously. Also, it can directly and indirectly prove the total effect among variables by taking measurement error into account. Finally, it can conduct an overall confirmatory factor analysis[1] using the latent variables. In other words, the structural equation model can present various results through other models such as the competitive model and the modification model.

Applying construction factors, the model is as follows Figure 1:

While the structural equation model shows the latent variables such as floor area, common area, and household area with a latent variable model, the observable variables have been shown with a measurement model, thus describing the causal relationship between the two models.

3.2 Principle Component Analysis

To extract the variables, a Principal Component Analysis (PCA) was performed. PCA is a method that finds principle components with correlated variables by using the structure that can explain the correlation among certain variables. This study conducted PCA in order to extract factors that have an eigenvalue larger than 1. Here the eigenvalue is an aggregate of all the variances among the factors that can be explained with each variable, and it is calculated by adding all the squared factor loadings of each factor of each variable. In other words, it is the ratio that indicates how much each factor is affected by each variable. The varimax rotation method was also used in order to achieve a clear and simple explanation among the factors. In the case of a factor loading that shows the level of correlation among factors, if the value is higher than 0.4, it is considered to be efficient, and if the value is higher than 0.5, it is considered to be an important variable (Roh, 2005)..

4. PREDOMINANT DRIVERS FOR COST ESTIAMTING

4.1 Cost Influence factors in Cost Estimating

Cost influence factors can be generally defined as the parameters that are used to derive the function, criterion, or rule used to estimate the cost of construction. Therefore, it is very important to define the influence factors and establish a criterion in order to accurately estimate construction costs.

(1) Cost per standard area - Quantity of material used (labor, concrete, size of mold, reinforcing iron, brick), employed number of personnel, cost per unit
(2) Regression model - Air, average number of stories, number of apartment houses, attached number of buildings, building dimensions, total area of construction, average size of housing, number of households, floor space index, configurative computation, ratio of underground construction, type

[1] Confirmatory factor analysis: A method in which research or theories is used to confirm a hypothetic model.

of floor, tiered/corridor/other, number of underground stories

(3) Cost weight theory - Area on lease, total building area, type of base foundation, number of stories, level of interior completion

(4) Monte-Carlo simulation - Air, configurative computation (the ratio between the floor area and the outer walls), average area per household (area, number of households), ratio of underground construction (area of the underground stories), average stories (building area), executed budget (planned budget), accomplished amount (actual investment), dimensions, building area

(5) ANN theory - Location, pottage, rebuilding, foundation work, duration of construction, number of apartments, type of management, average number of stories, underground parking lot, average size per household, total construction area, total number of households, completion

(6) Genetic Algorithm - Total area of construction, total area of parking space, building location, highest story, number of underground stories, number of apartments, number of households, types of apartment by size (peon), duration of construction, type of construction, location, type of project (redevelopment, rebuilding), base type (file, mat), completion, actual gains

(7) CBR - Total construction area, highest story, average number of households per story, total number of households, average size of house (peon), completion, air, total cost of construction, area, type of roofing, base type, type of underground stories

(8) DB - Total construction area, number of apartments, costs, superstructure work, costs of foundation work, cost of file, toilets, completion materials, costs of attached facilities, stairways, number of stories, type of tower, piloti, additional construction costs

4.2 Selecting the Predominant Influence factors

Influence factors are the standards related to cost estimation, and the particular ways in which factors are applied differ from case to case. However, commonalities can be discovered among the influence factors identified by previous studies. After analyzing the influence factors of previous cost estimation strategies, the variables are mainly divided into quantitative and qualitative variables, and the quantitative variables can again be divided into figurative variables and variables that require value judgment based on size and type. Based on the influence factors that have already been noted, and comparing these with real data and the characteristics of the factors, the following are the influence factors of cost estimation that can be applied in the reasoning phase of the CBR method:

(1) Quantitative influence factors: type of size (pyeong), number of households, floor space / total building area, construction area / building to land ratio, composition of households, number of elevators, number of stories, height between stories, number of underground stories, height between underground stories, duration of construction, ratio of underground construction, and other

(2) Qualitative influence factors: structure, type of roofing, piloti, the plan, and finish work.

4.3 Verifying the Predominant Influence factors

This section of the study will use PCA to interrogate the validity and grouping of the defined factors. Because there is a high correlation between the defined factors in this research, the ratio of household composition and the ratio of common use have been adopted as variables.

Table 2 Influence Factors' Factor Variable

Factors	Component			eigen value	Common variance
	1	2	3		
No. stories	0.922	0.630	0.021	2.929	27.943
height between stories	0.925	0.629	0.028		
Piloti	0.455	0.043	0.166		
composition of households	0.388	0.693	0.471	2.782	21.817 (49.760)
common use %	0.191	0.553	0.501		
No.ele	0.308	0.502	0.270		
household composition %	0.137	0.071	0.191		
No. households	0.275	0.239	0.800	2.049	21.576 (79.413)
size	0.107	0.647	0.665		
Construction area	0.115	0.052	0.506		

PCA was conducted on 10 of the cost influence factors. Among these factors, 3 factors that have a large amount of explanatory power have been extracted, follow Table2. A commonality can be found among these 3 factors, and the variance among them is 79.413%. In the social science fields, 60% is generally regarded as the standard variance for deciding on factors, and commonalities are regarded to have a component matrix value above 0.5 (Roh, 2005). This paper has excluded factors that have low correlation coefficient values.

For instance, there is a high correlation between the No. of stories and height of building because the numerical value of their correlation has contiguously been one. In general:

Total height = No. stories * height between stories

Furthermore, because of the strong influence between these two factors, in this case, they are assumed to be the same variable in statistics. In this research, the No. of stories and the piloti are set as the factors that influence structure.

The Predominant Drivers for Cost Estimation are as follows Table 3.

Table 3 Predominant Drivers for Cost Estimation

Group 1	Group 2	Group 3
number of stories with same height between stories	composition of households	number of households
	ratio of common use	size
Piloti	number of elevators	area of construction

5. ATTRIBUTE WEIGHT OF INFLUENCE FACTORS

5.1 Predominant Cost Drivers for the SEM

The factors that influence construction cost consist of independent and dependent variables, and a change in one variable leads to changes in other variables and in the overall result (e.g., construction cost). With this in mind, this study will attempt to conduct a confirmatory factor analysis on the selected influence factors using the observable and latent variables as models. Also, through a confirmatory factor analysis and exploratory factor analysis, this study will analyze the connection between the latent and observable variables. This verification process will utilize a covariance structure—which can clearly show the connections between multiple latent variables—because it seems to be the most suitable approach.

Moreover, the Amos program will be used to solve the problems that arise in the CBR reasoning phase. In other words, in order to enhance the accuracy of cost estimation, the Amos program is utilized to help solve the attribute weight problems that occur during the process of extracting and modifying data from past similar experiences.

Here, the pivotal factors are the cost estimation influence factors, and the cost estimating method is the cost-based approach. In the structural equation model, there are endogenous, exogenous, and disturbance variables. While exogenous variables are also known as independent error variables, endogenous and disturbance variables are also called dependent error variables. Here, exogenous variables are also defined as standard capacity variables, whereas endogenous variables are defined as normal variables. The influence factors used in this study are reviewed below:

1) Quantitative influence factors: number of stories, ratio of common use, composition of households, number of elevators, size (pyeong), number of households, area of construction

2) Overall influence factors: structure, type of roofing , the plan

A main objective of this study is to assume the cost influence factors from the perspective of a cost estimator. While Flanagan (1997) has defined the factors that influence the cost of any construction project as quantity, quality, and price level, Goule has defined them as project size, quality, location, duration of construction, and market condition.

Furthermore, other previous studies apply cost influence factors in a similar manner, and after examining these studies, it can be observed that quantity and price are the two factors that influence construction cost the most. This research is based on one of these cost estimating methods, which is known as the cost-based approach.

The influence variables can be placed into 3 variable groups. Piloti and number of stories can be grouped as *structure*; the ratio of common use, household composition, and the number of elevators can be grouped as *shape*; and the size of each household, the number of households, and the area of construction can be grouped as *size*. Here, structure, shape, and size are the variable groups or latent variables.

5.2 Applying the Cost Drivers in the SEM

5.2.1. Relationship between Influence factors and Cost Drivers

In this study, the CBR method is used together with all of the defined variables to solve the problem of determining attribute weights and to enhance cost estimation. The supplemented method is as follows Figure 2.

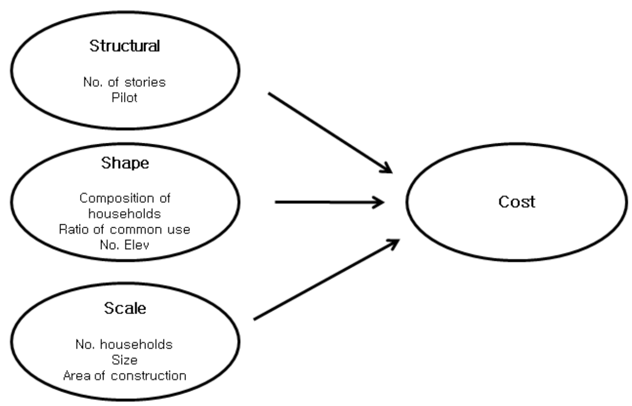

Figure 2 Relationship between Influence factors and Cost Drivers

With the proposed model, 2 major hypotheses can be made:

Hypothesis 1) The influence variables largely consist of structure, shape, and size.

Hypothesis 2) Structure, shape, and size will respectively have visible effects on cost.

5.2.2. Reliability Analysis of the Cost Drivers Model

Based on past data, this study analyzed the relation between the influence factors and cost of construction.

Two major characteristics that are essential to any data analysis are reliability and validity. Reliability depends on the accuracy of calculations and validity depends on the degree of adequacy in the data in relation to the purpose of that data (Roh, 2006).

Therefore, in order to determine the adequacy of the above variables, a reliability analysis was conducted by calculating Cronbach's alpha derivative.

Cronbach's alpha can be written as a function of the number of the test items and the average inter-correlation among the items. Below, for conceptual purposes, is the formula for the standardized Cronbach's alpha: Eq. (1)

$$\alpha = \frac{N \cdot \bar{c}}{(\bar{v} + (N-1) \cdot \bar{c})} \quad (1)$$

Here, N is equal to the number of items, c-bar is the average inter-item covariance among the items, and v-bar equals the average variance.

Table 4 Reliability coefficient for Influence factors

Content	Cost Drivers	influence factor	Cron α
Influence factor	Structural	Piloti	0.86
		No. stories	
	Shape	composition of households	0.74
		common use %	
		No. elevators	
	Size	No. households	0.78
		size(pyeong)	
		area of construction	
Cost	Construction		0.89

Cronbach's alpha derivative can be used to determine the consistency of the extracted variables, and it can be said that a certain factor is relevant if the value is higher than 0.5 (Wolf, 1992). As shown in the above table, the influence factors and quantity all indicate a Cronbach's alpha derivative above 0.5, thus demonstrating a high degree of reliability.

5.2.3 Determining the Attribute Weights

As the purpose of this paper is to determine what the cost influence factors consist of, and how these factors affect construction cost, an equation is derived that explains the correlation between the attribute weight applied factors and construction cost.

To achieve this goal, Confirmatory Factor Analysis (CFA)—which is useful in statistically verifying discriminant validation and convergence validation of items—was used to verify items that have undergone a 1st hand analysis. Furthermore, using AMOS 7, this study analyzed precisely how appropriate the observed variables and the model are, first conducting CFA on the parts of the model and then on the model as a whole.

To evaluate the goodness-of-fit, in order to calculate the best state of each of the factors in each of the phases, GFI (value above 0.9 ideal), AGFI (value above 0.9 ideal), X2 (smaller the better), and the value of P in relation to X2 (smaller than 0.05 ideal) were all referred to. The shape and size variables have 3 sub-items each, and because the goodness-of-fit converges to 1, these two variables were assessed as fitting the criterion.

Next, correlation analysis was conducted in order to evaluate the validity of the observed model, and it was proven that the model has discriminant validity among the variables that comprise it.

To confirm the validity of this study, it must be verified that the proposed structural equation model appropriately satisfies the first assumption. Validity evaluation of certain models is based on the Absolute Fit Measures, Incremental Fit Measures, and Parsimonious Fit Measures. Absolute Fit Measures rely on the index X2 and test the completeness of a certain model; this means it confirms the null hypothesis of the data and model at hand by emphasizing the overall goodness-of-fit. If the significance probability (C) is larger than 0.05(C>0.05), the null hypothesis is accepted and the model is evaluated as fit for the population of interest.

On the other hand, Incremental Fit Measures look for the concordant level between the proposed model and other pre-existing models using the Normed Fit Index (NFI). If the index is higher than 0.9, it is assumed that the model fulfils the goodness-of-fit.

This research has evaluated the overall appropriateness of the model based mainly on the GFI, CFI, NFI, X^2, and p figures. The results of this evaluation are as follows Table 5.

Table 5 Results of the Cost Drivers Model Validation

Classification	Result	Criteria
X^2 related to p	0.06	above 0.05 ideal
GFI : Goodness of Fit Index	0.736	>0.90 ideal
NFI : normed fit index	0.837	>0.90 ideal
CFI : Comparative fit index	0.852	>0.90 ideal

According to all the values gained from the validation process shown in the table above, all the numbers were in accordance to the various criteria. Thus, it is possible to assume that the model is appropriate.

The process of determining attribute weights is as seen in Figure 3 and is conducted with the Eq2~11.

Structural is represented by 2 influence factors.

$$X1 = P11 * Y1 + error \quad (2)$$
$$X2 = P21 * Y1 + error \quad (3)$$

Here X is equal to the influence factor, P is the critical ratio, Y is the Cost Driver, and error equals the error variable.

Shape is represented by 3 influence factors.

$$X'1 = P12 * Y2 + error \quad (4)$$
$$X'2 = P22 * Y2 + error \quad (5)$$
$$X'3 = P32 * Y2 + error \quad (6)$$

Scale is represented by 3 influence factors.

$$X''1=P13*Y3+error \quad (7)$$
$$X''2 = P23 * Y3 + error \quad (8)$$
$$X''3 = P33 * Y3 + error \quad (9)$$

Cost is affected by the 3 Cost Divers.

$$Z = P14 * Y1 + P24 * Y2 + P34 * Y3 + error \quad (10)$$

Here Z is equal to the construction cost, P is the critical ratio, Y is the Cost Driver, and error equals the error variable.

Here is the total structural equation:

$$Z = P14 * (X1 + X2) + P24 * (X'1 + X'2 + X'3) + P34 * (X''1 + X''2 + X''3)Y3 + \text{error} \quad (11)$$

It is also important to note that in this research, the critical ratio is the influence factors' attribute weights.

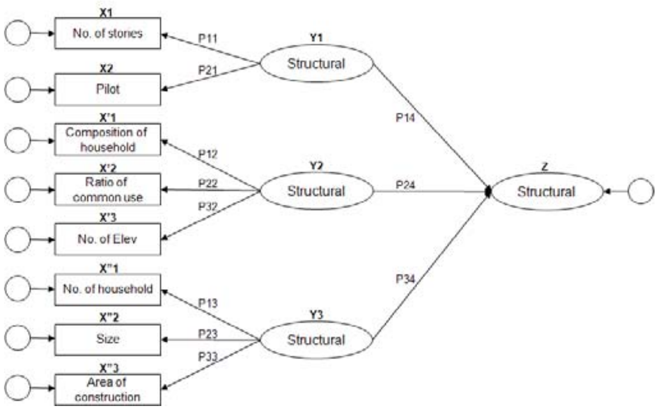

Figure 3 The early SEM for determining Attribute Weight

5.2.4 Verifying the Attribute Weights

Using the PCA process, Hypothesis 1—that the influence factors are largely comprised of 3 variables (i.e., Cost Drivers: structure, size, and quantity)—was verified in this chapter. Therefore, Hypothesis 1 can be reasonably accepted.

On the other hand, Hypothesis 2 regards the relationship between the selected influence factors and cost, and the grouping model results are shown in Table 6. As the critical ratio of all 3 cost drivers is larger than 1.96 (Roh, 2005), it can consequently be said that these factors effectively impact cost. Thus, Hypothesis 2 can also be accepted.

Throughout the course of this study, 3 variables or cost drivers-structure, shape, and size-have been identified as factors that affect construction cost. It has also been proven that in the measurement model, structure affects cost most significantly, then shape, then size. The findings so far can be illustrated as seen in the following figure

.

Table 6 Results of the Cost Drivers Model

hypothesis		correlation	path coefficient	critical ratio	Result
hypothesis 1		Consistence of variables	-	-	accept
hypos'	2-1	Structural => Cost	0.400	5.183	accept
	2-2	Shape => Cost	0.203	2.621	accept
	2-3	Size => Cost	0.326	4.029	accept

The influence factors and cost represented by the following Eq.12:

$$\text{Cost} = 0.400Y1 + 0.203Y2 + 0.328Y3 \quad (12)$$

6. CASE STUDY

This case study was conducted on two different apartment housing projects; The following Table 7 shows the criteria that were used to conduct the case study.

Table 7 Characteristics of the Construction projects

Classification	Project A	Project B
Size(pyeong)	84	59
No. households	46	38
Area of construction	492.38	365.7
Structure	Brick	Brick
Composition of Households	4	4
No. elevators	1	2
Type of common area	Corridor	Stairway
No. stories	13	12
Height between stories	41.2	36.35
Type of Roofing	Flat roof	Sloped roof
Piloti	2 Stories (2 Households)	2 stories (4 Households)
Shape of plane	' ㄴ' shape	' ㅣ' shape

Matching the weights derived above, this study also extracted similar cases based on the information from the 102 K-projects which had already been completed.

3 Cost Drivers

This data was categorized by:

①Structure, which is in the order of number of stories & Piloti

②Shape, which is in the order of household composition, ratio of common area, number of elevators

③Size, which is in the order of number of households, size(pyeong), area of construction

Then, when the data regarding the determined weights from past experiences are applied to the equation derived in section Eq.12, the ultimate cost of construction can be calculated:

$$\text{Cost} = 0.400Y1 + 0.203Y2 + 0.328Y3 \quad (12)$$

Cost = 283.060.583 + 135.245.259 + 225,458,651
= 643,764,493

Cost = 167,861,650 + 90,189,778 + 141,656,553
= 399,707,981

From apartment project A, 3, 2, and 4 similar examples were extracted according to structure, shape, and size, respectively, whereas 2, 3, and 3 similar examples were extracted from the B apartment project, respectively. Here, the quantity of material used from each project was calculated by applying the weights derived earlier. Then, the cost of construction was calculated using the cost per unit. The following Table 8 shows the results.

Table 8 Case Study Results

(*1000 won)

	Project A	Project B
actual construction cost	677,765,458	419,654,125
cost by SEM	643,764,493	399,707,981
error	14,000,965	19,946,144
% of error	-5.02%	-4.75%

The results calculated using the proposed structural equation model show 5.02% and -4.75% of error, respectively, for project A and project B. This means that the results are considerably reliable. Furthermore, the calculated cost of construction exhibits the tendency to be smaller than the actual cost. This is because errors in calculation are inevitable due to the weight problem of the path coefficient applied in the structural equation described in 12, as well as the problem of ignoring the error variables in the Cost Drivers Model .

Ultimately, by taking weight into account and applying the error variable and factor modification, more effective and accurate cost estimation can be achieved.

7. CONCLUSIONS

The primary objective of this research was to solve the problem of determining attribute weights, which arises during the reasoning phase of CBR, with a scientific approach, and ultimately develop a more effective cost estimation method. The basic items of consideration were qualitative and quantitative influence factors. However, qualitative influence factors need to be modified in order to be considered in the cost estimation process. Therefore, this research applied statistical methods to identify and extract 3 Predominant Cost Drivers: structure, shape, and size. Then, these quantitative factors were related to the qualitative factors to develop a Cost Drivers-structural equation model which can be used to estimate construction cost by considering attribute weight.

The major findings of this research are as follows:

(1) This research demonstrated the advantages of using a SEM to develop an effective CBR model for construction management. This methodology was developed to be particularly used in the initial phases of a construction project in which information is lacking. Based on CBR, the Cost Driver Model is a methodology that can be applied to schematic estimating in construction. CBR was used in this research to develop a prediction model which generates attribute weights through three optimization techniques, namely feature counting and GA.

(2) The Cost Driver Model was simulated using a SEM to provide a transparent and simplified representation. SEM is a comprehensive and flexible statistical method that can utilize a variety of statistical techniques simultaneously, such as regression analysis and correlation analysis. Furthermore, the SEM method analyzes data step by step; thus, it is easy to use the correlation model and to verify its results.

Additionally, by relating the predominant influence factors and the Cost Drivers model to determine attribute weight, this research was able to derive a correlation between grouped qualitative factors and cost. Not only can this Cost Drivers-structural equation model determine attribute weight and priority by using pair wise comparison, but it can also immediately assess the quantitative and qualitative influence factors and consider the measurement error of these factors. This could potentially assist construction personnel during crucial decision-making.

(3) While the validity of the traditional CBR method was confirmed through a review of previous studies, a case study verified the superior reliability (in terms of determining attribute weight) of the proposed model to the conventional CBR technique.

Also, using the proposed complementary method to estimate construction costs also reduces the margin of error yielded by the conventional CBR method. If this model is applied in practice (after further research and development), it could enhance decision-making and subsequently reduce the overall cost of a construction.

REFERENCES

[1] Aamodt, A., and Plaza, E. (1994) Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. AI Communications. IOS Press, Vol. 71, pp. 39-59.

[2] Arbuckle, J. L. (2006). Amos 7.0 User's Guide, SPSS Inc

[3] Bagozzi. R., and Yi. (1998). On the Evaluation of Structural Eequation, Jpurmal of Academy of Marketing Science, Vol. 16, No. 4, pp. 405-425

[4] Bode, J. (1998)Neural networks for cost Estimation. Cost Engineering, AACE Int., Vol. 40, No. 1, pp. 25-30

[5] Brown, C. H. (1983) Asymptotic comparision of missing data procedures for estimationg factor Loadings. Psychometrika, pp. 239-291

[6] Brown, R. L. (1994) Efficacy of the indirect approach for estimating structual equation models with missing data: A comparision of five methods. Structual Equation Modeling: A Multidisciplinaty Journal, 1, pp. 287-316

[7] Brown, M. W. (1984) Asymptotically distribution-free methods for the analysis convariance structure. British Journal of Mathematical and Statistical Psychology, 37, pp. 62-83

[8] Emory, C. W. and Cooper, D. R. (1991). Business Research Methods, 4th ed., Irwin

[9]Gu, C. W. (2007) A CBR-based Hybrid Model for Predicting Construction Duration and Cost based on Project Characteristics in Multi-Family Housing Projects

[10] Hegazy, Tarak. and Ayed, Amr. (1998) Neural Network Model for Paramatic Cost Estimation of Highway Projects, ASCE, Vol. 124, No. 3, pp. 210-218