Qualitative and Quantitative Cost Estimation: A Methodology Analysis

S. Aram[1], C. Eastman[2], J. Beetz[3]

[1]Digital Building Lab (DBL), College of Architecture, Georgia Institute of Technology, 723 Cherry St. NW Atlanta, GA, 30318; email: shiva_aram@gatech.edu
[2]College of Architecture and Computing, DBL Director, Georgia Institute of Technology, 723 Cherry St. NW Atlanta, GA, 30318; email: charles.eastman@coa.gatech.edu
[3]Department of the Built Environment, Eindhoven University of Technology, Den Dolech 2, 5800 MB Eindhoven, The Netherlands; email: j.beetz@tue.nl

ABSTRACT

This paper reports on the first part of an ongoing research with the goal of designing a framework and a knowledge-based system for 3D parametric model-based quantity take-off and cost estimation in the Architecture, Engineering and Construction (AEC) industry. The authors have studied and analyzed current cost estimation methods used in both the AEC and non-AEC industries in terms of their requirements, use contexts, methodologies, limitations and strengths to lay the groundwork for selecting the most suitable problem decomposition methods and cost estimation techniques to design a new framework. We have focused on determining the underlying methodologies of different cost estimation models and not just the techniques. Both qualitative methods used in early stages of design and quantitative methods used in more mature design stages are reviewed and their structure are analyzed.

INTRODUCTION

Efficiency, flexibility and accuracy of cost estimation methods significantly impacts every project, product development, and corporate success. Cost estimation is performed throughout a project and product development lifecycle and according to (AACE International 1997) can be categorized in five classes: concept screening, feasibility study, budget authorization and control, bidding/tendering, and check/control estimate. The major complexities of cost estimation are twofold: (i) the fact that at early stages of a project when quality of cost estimation has the highest impact on the success of a project and product outcome, there is limited information available; (ii) high variety of internal and external factors from design and engineering specifications to supply chain technologies and local regulations and limitations impact the total cost. Identifying all relevant factors, factoring them in the model, methodically defining their relationships with cost, and finally building a robust yet flexible and extendable cost model all add to the complexity of cost estimation.

This study is the first part of an ongoing research with the goal of designing a framework to define, quantify and retrieve cost driving properties of 3D parametric design models and to categorize and present them based on different criteria to users in the Architecture, Engineering and Construction (AEC) industry. In this study, the authors have analyzed current cost estimation techniques used in both the AEC and manufacturing industries. The analysis outcome is used to select the most suitable problem decomposition methods and cost estimation techniques for cost estimation in advanced design stages of construction projects. This in turn provides a stepping stone to design a framework for detailed quantity takeoff and cost estimation through extracting design model data and analyzing the extracted data.

Currently, we are working on other aspects of this framework including building a knowledge-based system through analyzing the domain knowledge regarding supply chain, structural and architectural design, and cost estimation, and developing rule based algorithms for data processing. These aspects of our framework will be explained in other papers currently under development.

COST ESTIMATION METHODS

Numerous studies have explored and implemented different cost estimation methods for generalized uses as well as specific use cases. We found many different implementations of qualitative methods used in the early design stages both in the AEC and manufacturing industries. Research efforts focused on the quantitative and analytical methods for later design stages have mostly targeted the manufacturing domain. Important reasons include the standardized production processes and higher consistency, reliability and generalizability of measurements, resource consumption, productivity rate and time and cost of each activity in a controlled manufacturing environment.

The AEC industry’s progress toward more standardization is accompanied with proliferation of two major trends of prefabrication and modularization. Many trades of the AEC industry and especially prefabrication sectors such as the precast concrete industry are increasingly using analytical cost estimation methods. The controlled production environment in construction prefabrication resembles that of the manufacturing industry. Thus, the lessons drawn from manufacturing including analytical cost estimation methods can provide useful insights for implementing them in areas like precast concrete which is the main focus area of this research effort.

Researchers have categorized cost estimation techniques in a variety of ways: (Cavalieri et al. 2004) classified cost estimation methods as analogy-based, parametric and engineering models. (Niazi et al. 2006) further divided intuitive methods into Case-Based Reasoning (CBR) and decision support systems, analogical methods into regression analysis and Artificial Neural Network (ANN), and analytical methods into breakdown, operation-, tolerance-, feature-, and activity-based cost modeling. (Chougule & Ravi 2006) classified cost estimation methods as intuitive, analogical, analytical, feature-based and parametric.

In both construction and manufacturing industries, the amount and level of detail of available design information at each stage of a project and the purpose of cost estimation determine the feasibility and suitability of the various cost estimation methodologies. Available information and cost estimation purpose are in turn dependent upon project phase and degree of design completion. Hence, the project phase provides a good basis for categorizing cost estimation research and methods.

INTUITIVE AND ANALOGICAL METHODS: EARLY DESIGN STAGES

Numerous studies have focused on conceptual design and initial design development stages of products and projects. Due to the lack of complete design information in early stages of a project, cost estimation solutions use qualitative methods in which new projects and products are compared to previous similar ones to identify the weight of different variables and degree of similarity in important aspects of projects, which are established by the researchers. As such, they are mostly categorized as analogical decision support systems (Niazi et al. 2006).

In response to limitations of traditional statistical techniques and to improve their performance in terms of accuracy and consistency, new techniques including the non-linear machine learning method of Artificial Neural Networks (ANNs), the problem-solving and learning method of Case-Based Reasoning (CBR), heuristic optimization algorithms like Genetic Algorithm (GA), and probability distribution optimization methods like Monte Carlo, and decision trees were introduced.

Two of the most frequently studied cost estimation methods for early design stages are ANN and CBR. The major advantages reported for ANN models are that they do not require the project cost to be defined as a specific function of cost-affecting variables. Also many studies in both construction and manufacturing have shown their higher accuracy compared to traditional regression models (Bode 2000; Kim et al. 2004; Cavalieri et al. 2004). Major advantages of CBR models are transparency of the process which turns it into a suitable decision support tool, the ability to handle missing attribute information from previous cases and the feasibility of long-term use due to ease of updating models through the addition of new cases (Arditi & Tokdemir 1999; Cavalieri et al. 2004)

The goal of these methods is to predict project costs with limited information provided in early stages of a project with an acceptable accuracy rate. These cost predictions are generally used by project owners for feasibility studies and budgeting purposes. While several different techniques are utilized for an early stage cost prediction, the applied methodologies are comparable in many aspects and can be generalized as the following steps:

- *Data collection* from previous sample projects of the same type and identification of important cost-driving attributes in the projects. These attributes and their values are used as inputs for the cost estimating system where the total project cost is the output. These are high level inputs. One example involves ten attributes of project type, scope, year, season, location, duration, size, capacity, water bodies, and soil condition which were used in a cost prediction study for highway projects (Hegazy & Ayed 1998). Another study (Tatari & Kucukvar 2011) collected values of 6 LEED certification categories in addition to building type, year and location data and used them as the system inputs to predict LEED certified projects' cost premium.
- *Identification and assignment of the optimal weights* to *input attributes* using different methods from linear statistical methods like Multiple Regression Analysis

(MRA) (Koo et al. 2010; Jin et al. 2012) to ANN (Günaydin & Doğan 2004; Tatari & Kucukvar 2011), GA (Kim et al. 2004) and decision trees (Doğan et al. 2008).

In these methods usually data from part of the collected project cases is used to train the model. The rest of collected project cases are used to test and validate performance of the built model in predicting total costs of projects, using the assigned weights for different attributes. The training involves systematically adjusting weights of attributes through comparing *predicted* output of the model – here the project cost – to the *actual* project cost. The goal is to minimize the error between predicted output of the model and the actual project output. One training method example is the back propagation algorithm which is the most broadly used method in ANNs. In this method, Mean Square Errors (MSE) are measured and minimized.

- *Prediction power assessment of the system*. Quality of a cost estimating model is evaluated by measuring its performance in predicting a project's cost using the final assigned weights for different attributes. As mentioned earlier, some of the collected project cases are used to compare predicted outputs of a model to the actual costs of those projects. Various algorithms and statistical methods can be used to assess the prediction power. For example, in the MRA method, the R^2, the coefficient of multiple determination, or the adjusted R^2 ($\bar{R}^2$) is used where the closer its value to 1, the higher the model's cost prediction accuracy. In the CBR method, different algorithms like the nearest-neighbor algorithm are used to calculate the similarity of the test project to training projects by a methodic comparison of their attributes. Finally the project case with the highest similarity rate is retrieved (An et al. 2007).

In the CBR method, the retrieved project is revised and adapted to the test project. Some CBR studies have applied subjective model revision approaches, while a few have used a systematic and analytical revision and adaptation process; one example is a construction CBR study that has applied a MRA-based process for revision (Jin et al. 2012). (Marzouk & Ahmed 2011) used four methods of null, weighted, neuro and fuzzy adaptation to revise the retrieved manufacturing cases.

Limitations of Intuitive and Analogical Methods. Part of the shortcomings of cost prediction methods stems from their inherent nature that inevitably rely on the availability of data from past similar projects. Methods like ANN can achieve more accurate results with fewer historical projects compared to traditional methods. Yet, they need a substantial number of similar historical projects with known project costs and cost driving attribute values (Bode 2000). This not only prompts feasibility issues due to rather scarce construction projects' data, but also hinders wide application of these methods because of the considerable time and funding needed to collect the required data. Better methods for reliable handling of incomplete historical data should be investigated (Kim et al 2012).

A few studies have tried to apply a systematic process to attribute selection. For example, (Marzouk & Ahmed 2011) conducted a statistical analysis on the results of a survey about cost driving factors in the pump station projects to identify the factors with the highest cost impact. While attributes selected for inputs of a cost estimation model significantly impact accuracy of predictions of the built models, most studies haven't analytically established that the selected attributes are the most critical cost driving factors of the selected test project. Often selected attributes were just a subset of what could be easily determined and collected from early stages of historical projects or were based on selections of previous studies.

Moreover, while the improved techniques that different cost estimation methods use to improve the accuracy of their cost prediction models most of the studies haven’t explored situations where the results are not satisfactory.

Shortcomings specific to each method have been determined and analyzed in numerous studies. Important examples are the difficulty in handling large numbers of variables (project attributes) and the requirement for establishing a cost function between inputs and outputs by regression analysis methods (Cavalieri et al. 2004). ANN models have been reported to require considerable time and effort to retrain and update when new cases are added, making them unsuitable for long-term use. Moreover, unknown relationships of inputs and outputs in the hidden layers result in a black box technique. Providing analytical explanations for the process and results to decision makers is thus difficult (Kim et al. 2004).

Furthermore, these methods and researches have not considered cost effects of technological changes such as process automation, prefabrication and Building Information Modeling (BIM). Other issues to be investigated include alternative contract types like design-build and IPD that allow concurrent design and construction, the selected structural, production and construction methods, and unusual design forms on their analogy and outcome.

ANALYTICAL METHODS: LATE DESIGN STAGES

Methods used in late design stages attempt to analyze a product design and its supply chain processes in detail to achieve more accurate cost estimation. As such, they can be categorized as quantitative or analytical methods and can be further divided into three categories of function-, feature- and activity-based cost estimation. Boundaries between these methods are blurred, and studies sometimes use a collection of cost factors associated with production processes, morphological design features, and consumed resources. Fig. 1 summarizes the methodology used by the analytical methods. Analytical methods at use a product decomposition structure and later need to integrate the collected knowledge about features, functions and activities. The analytical methods vary in terms of level of granularity present in their models.

An activity-based parametric solution for estimating cost of the foundry stage of disk brake production was developed in a study by (Qian & Ben-Arieh 2008). Major activities and their total cost of production were identified. Activities were divided into three categories: (i) activities with fixed costs in the batch level, (ii) activities with variable costs in the batch level and a linear relationship with the batch size, (iii) activities with diseconomy of scale. The major cost driver for activity *i* was defined (e.g. machining hour for the testing activity).
These parametric cost estimation studies have been mostly performed in the context of manufacturing industry and scope of studies has been typically limited to one part type and one phase of the production with limited parameters and activities.

In the study by (H’mida et al. 2006), the cost of manufacturing was estimated by modeling resources required for each activity and aggregating them to estimate the cost of operation process of features of the product. A product model describing the product from the manufacturing point of view was developed. The different available operations and alternate machine uses were identified for each feature. The cost reasoning model estimated the total cost as the sum of the manufacturing operations costs of all product features through solving a constraint satisfaction problem.

In the study by (Roy et al. 2008) to estimate cost of an automotive exhaust system production, the product was functionally decomposed, specification parameters describing each function were identified, historical data regarding processes and resource consumption rates were collected, and finally cost items were linked to each function to estimate cost of adding each function to the product.

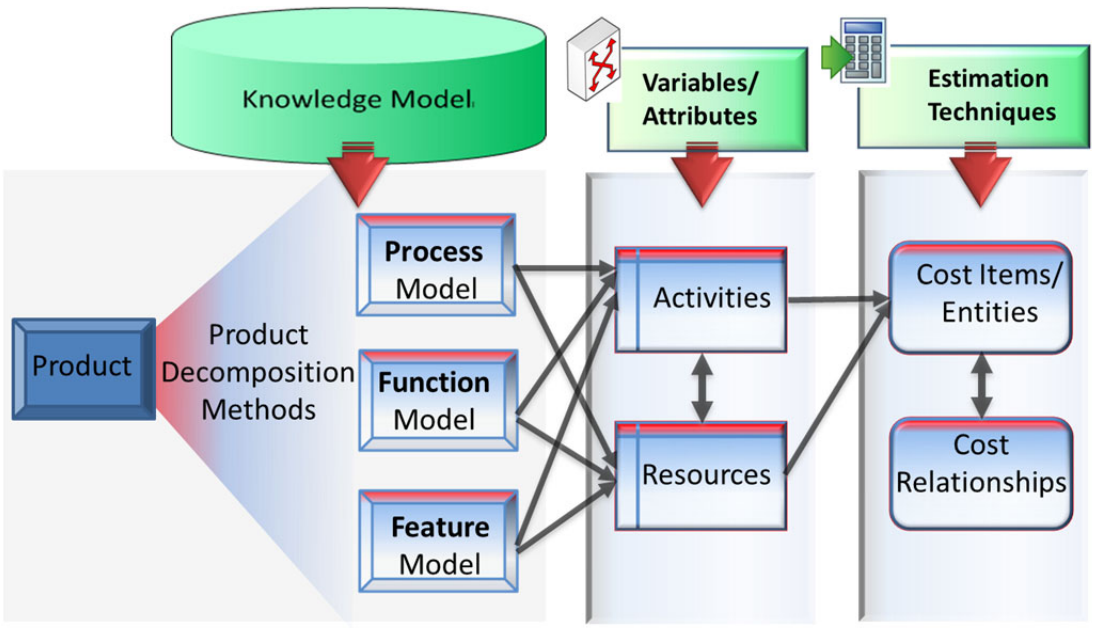

Figure 1: Integrated analytical cost estimation process

A study by (Chougule & Ravi 2006) created a system in which cost of activity resources were calculated using (i) various geometry, quality and production attributes of the product; (ii) a process model; and (iii) a 3D model for feeding and gating systems, as inputs of the cost model. Based on the reviewed research efforts on the analytical cost estimation the following methodology can be formulated:

- *Product Decomposition.* One of the product decomposition methods is selected. A product decomposition model for the standard product design is developed. Optional functions or features and alternative processes are defined. After an initial design, parts of the designed model that are of high complexity or of cost significance should be further broken down to achieve an appropriate level of detail.
- *Data collection.* Data regarding product, process, projects, functions, and cost driving parameters are collected from various resources including historical databases, engineering specifications, recording production supply chain, expert knowledge and judgments. This data will be used to identify cost driving parameters and their relationship with total cost of each activity, function or feature. Evaluating

the quality of acquired data to ensure of its measurability, reliability and completeness (Cavalieri et al. 2004) is important for defining accurate cost functions.

- *Cost driving parameters/variables/attributes are specified* for each unit of the decomposition model –i.e. each activity, function or feature – through analyzing the supply chain and eliciting knowledge of domain experts. For accurate cost estimation, selected attributes should reflect all aspects of a product's lifecycle. Various categories of parameters concerning geometry, quality, aesthetics requirements, engineering performance and production technology should be analyzed.
- *Define cost relationships/functions.* Cost behavior of units of the product decomposition model with regard to changes in the magnitude of those units is analyzed. These cost functions are expressed mathematically by equations between parameters defining each unit to total cost of the unit which basically requires a regression analysis. Usually an operation process involving several activities is required to produce a feature or provide a function. Hence, analyzing cost behavior of functions and features often leads to further decomposing them into activities.

In terms of cost relationships, activities most frequently belong to one of the four types of (1) fixed, (2) variable (proportionate to activity volume), (3) mixed (with a fixed and a variable cost portion), and (4) step (fixed within specific activity volume intervals , but jump to a higher level from one interval to the next) (Hartgraves & Morse 2012). In some cases activities have nonlinear and sometimes multi-variable cost relationships. In late design stages and in presence of the complete required design data and with a sufficient level of detail in decomposing a product, the cost behavior of activities can be adequately approximated by a linear function.

- *Aggregate cost relationships and estimates.* Aggregation of cost functions for all units of product decomposition model provides the total product cost. Cost aggregation can be done on various levels and each provides a unique insight into the product cost: (i) when a variable affects several different activities or features and hence is repeated in different cost functions, these functions can be aggregated to analyze the overall impact of each variable on the total product cost; (ii) all cost functions related to each specific resource can be aggregated throughout the supply chain to identify resources that comprise the largest portion of the overall cost; (iii) aggregation of the cost of activities at each stage or sub-process to focus on stages or processes with highest share of total costs or higher cost rates than industry averages.

CONCLUSION AND NEXT STEPS

This study has analyzed various qualitative and quantitative cost estimation methods in construction and manufacturing in terms of their requirements, methodology, limitations and strengths. The next step for the authors is to design a framework and a knowledge-based system that integrates the methods across cost estimation domains for defining, quantifying, querying and retrieving cost driving properties of the 3D parametric design models and categorize and present them based on different criteria.

REFERENCES

AACE International (1997). *"Recommended practice: cost estimate classification system – as applied in engineering, procurement and construction for the process industries."* SSVR 18R-97, Morgantown, WV.

Arditi, D., Tokdemir, O. B. (1999). *"Comparison of case-based reasoning and artificial neural networks."* J. Comput. Civil Eng. 13(2–3):162–169.

Bode J. (2000). *"Neural networks for cost estimation: simulations and pilot application."* Int. J. Prod. Res., 38(6), 1231–1254.

Cavalieri, S., Maccarrone, P., Pinto, R. (2004). *"Parametric vs. neural network models for the estimation of production costs: A case study in the automotive industry."* Int. J. Prod. Econ. 91,165–177.

Chougule, R. G., and Ravi, B. (2006). *"Casting cost estimation in an integrated product and process design environment."* Int. J. of Comp. Integ. M. 19(7), 676-688.

Doğan, S. Z., Arditi, D., and Günaydin, H. M. (2008). *"Using decision trees for determining attribute weights in a case-based model of early cost prediction."* J. Constr. Eng. M. ASCE, 134(2), 146–152.

Günaydin, H. M., and Doğan, S. Z. (2004). *"A neural network approach for early cost estimation of structural systems of buildings."* Int. J. Proj. Manag. 22, 595–602.

Hartgraves, A. L., and Morse, W. M. (2012). *"Chapter 2: Cost behavior, activity analysis, and cost estimation." Managerial Accounting*, 6th ed., Cambridge Business Publishers, Westmont, IL.

Hegazy, T., and Ayed, A. (1998). *"Neural network model for parametric cost estimation of highway projects."* J. Constr. Eng. M. ASCE, 124(3), 210–218.

H'Mida, F., Martin, P., and Vernadat, F. (2006). *"Cost estimation in mechanical production: The cost entity approach applied to integrated product engineering."* Int. J. Prod. Econ. 103, 17–35.

Jin, R., Cho, K., Hyun, C., and Son, M. (2012). *"MRA-based revised CBR model for cost prediction in the early stage of construction projects"* Expert Syst. Appl. 39 (5), 5214–5222.

Kim, G. H., Yoon, J. E., An, S.H., Cho, H. H., and Kang, K. I. (2004). *"Neural network model incorporating a genetic algorithm in estimating construction costs."* Build. Environ. 39, 1333–1340.

Kim, G. H., An, S.H., and Kang, K. I. (2004). *"Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning."* Build. Environ. 39, 1235–1242.

Koo, C., Hong, T., Hyun, C., and Koo, K. (2010). *"A CBR-based hybrid model for predicting a construction duration and cost based on project characteristics in multi-family housing projects."* Can. J. Civ. Eng. 37, 739–752.

Marzouk M. M., Ahmed, R. M. *"A case-based reasoning approach for estimating the costs of pump station projects."* J. Adv. Res. 2 (4), 289-295.

Niazi, A., Dai, J.S., Balabani, S., and Seneviratne, L. (2006). *"Product cost estimation: technique classification and methodology review."* J. Manu. Sci. E-T ASME, 128 (2), 563–575.

Qiana, L. and Ben-Arieh, D. (2008). *"Parametric cost estimation based on activity-based costing: A case study for design and development of rotational parts."* Int. J. Prod. Econ. 113, 805–818.

Roy, R., Souchoroukov, P., and Griggs, T. (2008). *"Function-based cost estimating."* Int. J. Prod. Res. 46(10), 2621-2650.

Tatari, O., and Kucukvar, M. (2011). "*Cost premium prediction of certified green buildings: A neural network approach."* Build. Environ. 46, 1081–1086.