Adding Qualitative Context Factors to Analogy Estimating of Construction Projects

Paula Riquelme*, Alfredo Serpell

Department of Construction Engineering and Management, Pontificia Universidad Católica de Chile, Vícuña Mackenna 4860, Santiago, 7820436, Chile

Abstract

Existing estimating models have certain shortcomings in the management of historical data. There is a need of defining more objective and consistent criteria for the selection of historical construction data to be used for estimating. In this perspective, a methodology based on historical information, which incorporates qualitative context factors to the structure and use of this information for cost estimating, such as project complexity, environmental conditions and characteristics of workmanship, among others, is proposed. A list of qualitative project context factors that are most influential for construction projects’ cost and productivity is presented. Additionally, a context model that includes these variables is described together with an explanation of how they are incorporated into the cost estimate. It is concluded that the incorporation of qualitative context factors in cost estimating improves the use of historical information and that most critical aspects to achieve this feature are the creation of a reliable site-work feedback system and the correct structure of historical information.

© 2013 The Authors. Published by Elsevier Ltd. Selection and/or peer-review under responsibility of IPMA

Keywords: Cost estimating; critical cost factors; historical information; context

1. Introduction

Despite decades of effort to improve the accuracy of the estimates, large infrastructure projects continue to be affected by delays and cost overruns (Liu & Napier, 2010), which are not limited to an owner or a particular type of projects (Shane et al., 2009), nor only to complex projects (Baloi & Price, 2003). Cost estimate is considered one of the most critical and important stages of a construction project
(Jrade & Alkass, 2007), and many studies have been conducted to determine the causes of these differences and suggest improvements based on the use and structuring of historical information (Jrade & Alkass, 2007; Lee, 2008; Chou, 2009; Figueiredo & Philipenko, 2010; Honsinger et al., 2010).

However, existing estimating models have shortcomings in the management of available historical data. The cost and the justification of the time and resources to support the process of collecting, reviewing and organizing data in a project can be a significant challenge for many organizations (Figueiredo & Philipenko, 2010), so eventually the decisions are taken based on the criteria and experience of each estimator. Therefore, there is the need of defining objective and consistent criteria for the selection of appropriate historical construction data to be used for cost estimating, such as construction productivity. Following this, it is proposed a methodology based on historical information, which structures and incorporates qualitative context factors such as complexity, soil conditions and characteristics of the workforce of the project, etc.

It is presented a list of qualitative context factors and sub factors that are most influential in construction projects’ cost. Using these factors and sub factors, it is proposed a model to define the context of a construction project. Finally, is described an initial approximation to a method that integrates the context model proposed and a criterion for determining similarity between projects, into the cost estimating of construction projects.

2. Background

Studies indicate that repeatedly the same problems have caused cost overruns of different projects and that much knowledge can be learned from studying the past (Shane et al., 2009). Therefore, the knowledge and individual experience of people and historical examples should be stored, structured and continuously fed back for easy retrieval and analysis by others (Baloi & Price, 2003).

The estimation by analogy is a cost estimation method commonly applied which is based on being able to determine historical cost of activities or items and use them as reference to predict costs of new activities or items proposed (Greves & Joumier, 2003). Therefore, a primary consideration for its use is to be able to determine the differences between current and past items or activities to estimate (Greves & Joumier, 2003), because from this will depend the required adjustments. This process of comparison between a construction project and another, becomes complex when considering the incredible degree of customization and flexibility that the industry offers to the consumer (very few products are so targeted to meet the desires of a particular consumer as construction projects) (Sawhney et al., 2004).

This situation occurs in a similar way when estimators reuse past activities yields, to estimate the yield of new activities, because fluctuations occur mainly due to different conditions on which the specific activity was performed (Kisiltas & Akinci, 2009). In this case, the contexts of projects’ realisation are compared.

Many authors propose a list of variables that affect the productivity and costs of construction projects. However, studies indicate that time and cost prediction techniques used in the construction industry have shortcomings regarding the lack of incorporation of qualitative variables to the estimation process and the lack of definition of criteria for the selection of the relevant historical data (Table 1 and Table 2 respectively). This may be explained because qualitative aspects are difficult to evaluate (Elhag & Boussabaine, 1997) and because the time and resources to support the process of collecting, reviewing and organizing the project data can be a significant challenge for many organizations (Philipenko & Figueiredo, 2010).
3. The problem

The quantitative knowledge is defined as cost elements and known structures of elements which form the basis of a cost estimate and are measurable (such as number of parts or types) (Rush & Roy, 2001). The qualitative knowledge is defined as the assumptions and judgments that cost estimators and engineers do during the generation of an estimate. These assumptions and judgments are related to how an estimator refers to past projects as a basis for the generation of a new estimate (Rush & Roy, 2001). This qualitative knowledge is used for choosing or adjusting multiple values of past performance according to the execution characteristics of the new project (context). The problem is that cost prediction techniques in the construction industry only consider relevant factors that can be quantified (Elhag et al., 2005). Since most of the critical factors that affect project costs are of qualitative nature, estimators do not have all the relevant information related to past projects, not have a defined or applied selection method or a formal set of these, and for these reasons, they carry out a subjective assessment of the observed situation (Kisiltas & Akinci, 2009).

The research question was stated as: how to determine, in a consistent and objective way, the performance to be used for activities under different conditions or contexts, on the basis of historical construction projects’ data?

The main question generated the following sub-questions:

- What are the main contextual factors that create variation in the cost or productivity of construction projects?
- How to incorporate the qualitative context in the development of cost estimates for these projects?

Table 1. Methodologies for estimating by analogy and its main shortcomings

<table>
<thead>
<tr>
<th>Author</th>
<th>Description of the estimation methodology</th>
<th>Shortcoming in relation to the selection of relevant information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proverbs et al. (1999)</td>
<td>Method for estimating labour requirements and costs for international projects. Is based on determining labour productivity rates per area for concrete activities on site.</td>
<td>Productivity differentiation criteria only based on each country surveyed.</td>
</tr>
<tr>
<td>Yu (2006)</td>
<td>Model for conceptual cost estimation of construction projects (PIREM) that integrates several methods. The main cost items are modelled by a nonlinear parametric function.</td>
<td>Key parameters of the cost equations are only aspects of the project as superstructure type, foundation type, length of the stack, and so on.</td>
</tr>
<tr>
<td>Huawang, &amp; Wanqing (2008)</td>
<td>Costing methodology for construction projects which integrates rough set (RS) and Artificial Neural Network (ANN). RS is applied to find the relevant factors of cost, to be used as inputs in an ANN to predict the cost of a project.</td>
<td>The models are made mostly with quantitative variables of project design: total height, area, type of structure, level of project management, period, underground area and level of project management.</td>
</tr>
<tr>
<td>Chou (2009)</td>
<td>Web-based CBR system applied to early cost budgeting for pavement maintenance project. Compare the historical data, item-level work through a library of cases.</td>
<td>The similarity between projects is based on quantitative factors, mostly, except project location and terrain (flat, hill or mountain). Applicable to a single type of project.</td>
</tr>
<tr>
<td>Rush &amp; Roy, 2001</td>
<td>Model to estimate project costs through the use of expert judgment and analogy. Model the steps that make up the reasoning process used by an expert to make a cost estimate.</td>
<td>Indicates that the analogy is based on similar projects, which lists aspects that are usually used, but does not specify the most relevant and how to determine the similarity. Authors recommend research on how to make comparisons between projects.</td>
</tr>
<tr>
<td>Mohamed &amp; Celik, (2002)</td>
<td>Knowledge-based integrated system for cost estimates and construction programs. Users enter general information about the project, select types of materials and evaluate productivity factors indicated.</td>
<td>With these productivity factors related to the project context, it affects only the duration of the project. The estimated cost only depends on design parameters, quantities, unit prices and materials.</td>
</tr>
</tbody>
</table>
The postulated research hypothesis is that it is formally possible to incorporate qualitative aspects of context in estimating the cost of a construction project, based on the structuring and use of historical information.

Table 2. Database for cost estimating and its main shortcomings

<table>
<thead>
<tr>
<th>Author</th>
<th>Database description</th>
<th>Shortcoming identified</th>
</tr>
</thead>
</table>
| Kiziltas & Akinci, (2009) | Sets parameters must contain a database of historical project costs (design, construction process, the construction site and the project). | 1) Factors 4 focused in detail on specific activities.  
2) Does not indicate how to determine the similarity based on the factors.  
3) It shows how to integrate the database with an estimation methodology. |
| Honsinger, 2010          | Important aspects Guide to generate a database focused on parameter estimation and control. Relevant parameters are indicated similarity: scope, size, year, key assumptions and conditions. | 1) Factors quite general and with no explanation of your choice.  
2) In determining similarity indicates an example of formula to filter, but does not define fixed criteria.  
3) It shows how to integrate the database with a system of estimation. |

4. Methodology

The methodology used was to carry out a comprehensive literature review initially on the following topics: a) cost estimation methodologies based on the use of historical information, b) factors that affect the performance and productivity, and that cause overruns or inaccuracy in construction projects, c) methods for determining the similarity between projects. Through an arrangement, classification and application of the criterion of affinity, a list of 33 factors of 26 authors was obtained. These factors are shown in Figure 1.

Fig. 1. Factors affecting the cost and productivity of construction projects
Subsequently, and with the aim of focusing the study on the most relevant factors, a frequency study by author and a Pareto analysis were applied, obtaining 6 main factors in this way. To validate these factors, semi-structured interviews with experts in the area of cost estimation of construction projects of engineering and construction Chilean firms were conducted. Only 5 principal factors, described in Figure 2 were selected by these 9 experts from 9 companies (institutional or residential building, housing, installation of industrial plants, etc.). In addition, a list of the main sub factors to consider for each factor and the variables that determine the possible conditions was obtained.

With qualitative context factors and sub factors, and the possible conditions for each one, a Context Model for a construction project is proposed. With the Context Model and similarity criteria for projects, a methodology to incorporate the qualitative factors of context to cost estimation was defined. As future work, a case study will be conducted to validate the results obtained by applying the proposed methodology.

5. Context Model

The proposed Context Model was created to define the context in which a construction project develops. The model is defined by a matrix of factors and sub factors and the conditions associated with each one of the sub factors (see Table 3). The 15 context sub factors are qualitative variables, and their values are the definitions of possible conditions. These definitions were built in base to the information gathered from interviews. The purpose is to provide information about a context that it is too complex normally or ill-defined with a conventional description in quantitative terms (Baloi & Price, 2003).

To define the context of a project (historical or to be estimated), the estimator must select the alternative condition for each sub factor that best represents what happened or what is expected to happen. Every alternative condition has a positive integer associated with it, necessary to calculate the similarity. For each sub factor, conditions are ordered for most to least favourable and numbers are assigned starting at 1.

To better visualize this, two examples of variables and alternative definitions of condition are performed for two contextual sub-factors (see Table 4 and Table 5).
6. Methodology for incorporating qualitative factors to the estimation of costs

The proposed methodology defines how to structure the historical project information, how to select the most relevant historical project information to the estimation of the new project and finally, how to apply this information to estimate the new construction activities yields. This methodology is based on: 1) the structure of the historical project information and 2) the processing and delivery of information.

Table 3. Generic structure of proposed Context Model

<table>
<thead>
<tr>
<th>Context factor</th>
<th>Context sub factor</th>
<th>Definition of condition</th>
<th>Construction projects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>(P_1)</td>
</tr>
<tr>
<td>(f_i)</td>
<td>(sf_{i1})</td>
<td>(\text{def. cond.}_{i11})</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\text{def. cond.}_{i12})</td>
<td>x</td>
</tr>
<tr>
<td>(\ldots)</td>
<td>(sf_{21})</td>
<td>(\text{def. cond.}_{21p})</td>
<td>x</td>
</tr>
<tr>
<td>(f_n)</td>
<td>(sf_{n1})</td>
<td>(\text{def. cond.}_{n11})</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>(\ldots)</td>
<td>(\ldots)</td>
<td>x</td>
</tr>
</tbody>
</table>

Table 4. Possible conditions for degree of innovation of the project

**FACTOR 1: PROJECT COMPLEXITY**

<table>
<thead>
<tr>
<th>Sub factor</th>
<th>Variable</th>
<th>Possible conditions (in order from worst to best)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree of innovation of the project</td>
<td>Innovation of design, materials, technology or implementation methodology of the project in relation to the company or the market in general</td>
<td>Project design, materials, technology or implementation methodology very different from those made by the company or the industry in general, so it's pretty difficult to get staff with the appropriate knowledge or reuse past experiences.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Project design, materials, technology and methodology with some novel features and similar to those made previously by the company or industry in general, it is necessary to search for experts only for some subjects, and past knowledge is available for reuse in multiple aspects.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Project design, materials, technology and implementation methodology quite similar to others conducted by the company or the industry in general, so there are many experts available and you can reuse past knowledge in most of the components of the work.</td>
</tr>
</tbody>
</table>

Table 5. Possible conditions for clarity of the specifications

**FACTOR 2: QUALITY OF PROJECT INFORMATION AND DOCUMENTATION**

<table>
<thead>
<tr>
<th>Sub factor</th>
<th>Variable</th>
<th>Possible conditions (in order from worst to best)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clarity of specifications</td>
<td>Degree of self-explanation and consistency</td>
<td>The specifications provided have inconsistencies, aspects not well defined or ambiguous information in many critical parts or components to the project.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The specifications provided have inconsistencies, aspects not well defined or ambiguous information in a few components or parts that are not critical to the project.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The specifications provided are self-explanatory and are consistent with each other, so it does not generate doubts or problems of interpretation.</td>
</tr>
</tbody>
</table>
6.1. Structuring of the historical project information

The organization must provide feedback to a database that comprises three main types of information:

- **Overview of the project**: project name, client, contract type, estimated and actual date of start and end, and location. The estimator can add the information as deemed necessary.
- **Background information for the project**: context associated with the execution of each project through the defined Context Model.
- **Activity-based costing information**: size of the activity and amount of resources required (both with their unit of measurement). This will determine the yields associated with resources and not with price, since they vary over time. The unit price of each resource must be entered each time the estimator should perform cost estimation.

On the other hand, the estimator should define settings to the proposed generic context model, in order to focus on the types of projects that are carried out by his/her company. These adjustments relate to:

- **Factors**: remove, add, or decomposed sub factors or contextual factors.
- **Conditions**: make clear the definition of possible conditions, add new conditions, and take out those which do not apply, or break the existing ones into more detail.

![Diagram](image.png)

Fig. 3. Application of the proposed cost estimating system

6.2. Processing and delivery of information

The processing and delivery of information is mainly based on the use of a similarity criterion between construction projects to define a method of historical selection of projects which are more alike to the project to be estimated. The adjustments required from the estimator are:

- Define the relative weights of context factors and sub-factors, necessary for the application of the criterion of similarity between projects.
- Carry out the design review and adjustments as required (contents and format).
- Define the minimum similarity value required for the project to estimate and historical projects (historic projects with lower overall similarity that the limit indicated, should not be considered for the estimation of costs).
6.3. Similarity criterion between projects

The determination of similarity between projects is based on the criteria used by Serpell (2010) and Chou (2009), where the overall similarity is built on a number of called local similarity functions, one for each attribute describing the project involved. Thus, the overall similarity between two cases is calculated as the weighted sum of local similarities.

For the proposed system, the Local Similarity (between sub factors) is determined first, then the Global Similarity (between factors) and finally, the Project Similarity (between projects).

- Local Similarity (at the level of contextual sub-factors):

\[
Local \ Sim \ (P_{ij}, Q_{ij}) = 1 - \frac{\text{dist}(V_{P_{ij}}, V_{Q_{ij}})}{R_{max}} \tag{1}
\]

where:

- \( P = \) Input project (to be estimated).
- \( Q = \) is a project of the historical database.
- \( i = \) is an individual sub factor that belongs to factor \( j \).
Global Similarity (at the level of contextual factors):

\[
Global Sim (P_j, Q_j) = \frac{\sum_{i=1}^{n} wsfi \times Local Sim (P_{ij}, Q_{ij})}{\sum_{i=1}^{n} wsfi}
\]  

where:
- \(P\) = Input project (to be estimated).
- \(Q\) = is a project of the historical database.
- \(n\) = the number of sub contextual factors that belong to the same factor \(j\).
- \(i\) = an individual sub factor from 1 to \(n\), which belong to factor \(j\).
- \(wsfi\) = the relative weight function of the sub factor \(i\).
- \(Global Sim\) = the similarity function between factors of projects.

Project Similarity (at the level of project):

\[
Project Sim (P, Q) = \frac{\sum_{i=1}^{n} wfi \times Global Sim (P_i, Q_i)}{\sum_{i=1}^{n} wfi}
\]  

where:
- \(P\) = Input project (to be estimated).
- \(Q\) = is a project of the historical database.
- \(n\) = the number of contextual factors (5 generics).
- \(i\) = is a single factor from 1 to \(n\).
- \(wfi\) = is the relative weight of the factor of context \(i\).
- \(Project Sim\) = the similarity function between projects.

To clarify this, it shows a small example of the calculation of the similarity between a project A and a project B, with the assumption that there are only two relevant context factors (Table 6 shows the required information).

Local Sim (\(A_{11}, B_{11}\)) = 1 - dist(\(V_{A_{11}}, V_{B_{11}}\))/\(R_{max}\) = 1 - Abs(\(V_{A_{11}}, V_{B_{11}}\))/\(R_{max}\) = 1 - (1-1)/(2-1)=100%

Local Sim (\(A_{12}, B_{12}\)) = 1 - dist(\(V_{A_{12}}, V_{B_{12}}\))/\(R_{max}\) = 1 - Abs(\(V_{A_{12}}, V_{B_{12}}\))/\(R_{max}\) = 1 - (3-1)/(3-1)=0%

\[\Rightarrow\] Global Sim (\(A_1, B_1\)) = \(n\) = \(\sum_{i=1}^{n} wsi\) * Local Sim (\(P_{ij}, Q_{ij}\))/\(\sum_{i=1}^{n} wsi\) = 0,1*100% + 0,9*0%=10%

Local Sim (\(A_{21}, B_{21}\)) = 1 - dist(\(V_{A_{21}}, V_{B_{21}}\))/\(R_{max}\) = 1 - Abs(\(V_{A_{21}}, V_{B_{21}}\))/\(R_{max}\) = 1 - (4-2)/(4-1)=33%

\[\Rightarrow\] Global Sim (\(A_2, B_2\)) = \(n\) = \(\sum_{i=1}^{n} wsi\) * Local Sim (\(P_{ij}, Q_{ij}\))/\(\sum_{i=1}^{n} wsi\) = 1*60%=60%

\[\Rightarrow\] Project Sim (\(A, B\)) = \(n\) = \(\sum_{i=1}^{n} wsi\) * Global Sim (\(P_i, Q_i\))/\(\sum_{i=1}^{n} wsi\) = 0,3*10% + 0,7*60% = 45%
Table 6. Context information for project A and project B.

<table>
<thead>
<tr>
<th>Context factor</th>
<th>Factor relative weight</th>
<th>Context sub factor</th>
<th>Sub factor relative weight</th>
<th>Definition of condition</th>
<th>Value of condition</th>
<th>Projects</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_1$</td>
<td>0.3</td>
<td>$sf_{11}$</td>
<td>0.1</td>
<td>def. cond.$_{111}$</td>
<td>1</td>
<td>$P_A$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>def. cond.$_{112}$</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>def. cond.$_{113}$</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$f_2$</td>
<td>0.7</td>
<td>$sf_{21}$</td>
<td>1</td>
<td>def. cond.$_{211}$</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>def. cond.$_{212}$</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>def. cond.$_{213}$</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>def. cond.$_{214}$</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

6.4. Reporting of information to estimator

The methodology recommends the development of two types of reports: 1) Reports required to estimate the costs of a new project, and 2) Reports for subsequent analysis at the end of a project.

- **To estimate yields:** Historical Projects Preliminary Report (historical projects in order of Project Similarity) and Yields Requested Activities Report (maximum, minimum and average yields on similar projects by activity, until the minimum Similarity value defined).
- **From these,** the estimator discusses and estimates the final yield of the activity for the new project, considering the following aspects: 1) Local and Global Similarity, indicating reasons and magnitude of adjustment in addition to the assessment context, which defines whether it is adjusting up or down, 2) Associated with the size of the activity and it is estimated. For a larger size of the activity, the average yield should be higher (learning by repetition).
- **For further analysis:** Report of comparison between the estimated and actual information (feedback for the Context Model and the proposed methodology).

7. Conclusion

This paper describes a study of the contextual factors that impact cost and performance of construction projects, understanding context as the characteristics of the situation in which projects are carried out. Throughout the study, the main achievement was in defining the main factors: project complexity, quality of project information, weather conditions, and characteristics of labor and site conditions. Through its utilization it was possible to define a context model for construction projects, which allows defining the context execution of the project, and a methodology to incorporate these qualitative factors for cost estimating.

The methodology defines how to structure, retrieve and use the available historical data, using an objective and consistent criteria to define similarity between contexts of construction projects. It is concluded that the incorporation of the qualitative context factors to the cost estimation process through the proposed system, could improve the application of historical information, as it allows: 1) to identify, understand and structure those contextual aspects that have a critical influence on yields and costs, allowing to define which specific information is required for projects, 2) to establish and improve the selection criteria for relevant information within an organization, which involves minimizing personal
biases in cost estimates, and 3) to maintain and increase the implicit knowledge of experts and past experience within the organization.

However, the most critical aspects for successful implementation of this methodology are the creation of a system of permanent monitoring, control and feedback of the results obtained, in addition to structuring and updating the historical data available (with new data and further adjustments required). The validation of the results obtained by applying the proposed system will be made in future studies, by an application to a company. For future research it is recommended to complement the Context Model with activity level factors in order to establish more detailed criteria of similarity.

References


