An investigation into the implementation of a predictive model for the early stage estimation of elemental building costs

Link to publication record in Manchester Research Explorer

Citation for published version (APA):

Published in:
Proceedings of the 3rd Scientific Conference on Project Management, "Clustering in Construction Project Management"

Citing this paper
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy
If you believe that this document breaches copyright please refer to the University of Manchester’s Takedown Procedures [http://man.ac.uk/04Y6Bo] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.
This is the author’s version of a work that was submitted/accepted for publication in the following source:


This file was downloaded from: https://www.escholar.manchester.ac.uk

© Copyright 2004 PM-03, J P Pantouvakis

Reproduced in accordance with the copyright policy of the publisher. **Notice:** Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source.
AN INVESTIGATION INTO THE IMPLEMENTATION OF A PREDICTIVE MODEL FOR THE EARLY STAGE ESTIMATION OF ELEMENTAL BUILDING COSTS

Michail Soutos, David Lowe
Project Management Division, Manchester Centre for Civil & Construction Engineering, UMIST, Manchester, UK
M.Soutos@postgrad.umist.ac.uk

ABSTRACT

Recent research at UMIST has resulted in the production of “ProCost”, a piece of software that predicts the final price of a proposed building at an early stage.

This paper analyses the results of a nationwide questionnaire survey, as well as the outcomes of a series of interviews with three major UK based Quantity Surveying firms. The results of this research show that the vast majority of cost estimators believe that a single figure cost estimate is insufficient to meet their forecasting needs. For this reason, different formats for breaking down building costs into sub-elements are investigated.

Based on the chosen format, regression analysis is used to develop a series of predictive models. One of these models is discussed in the current paper and the values predicted by it are compared with actual values. An analysis of the variance between the two follows.

KEYWORDS

Cost estimation, elemental cost analysis, neural networks, regression analysis,

INTRODUCTION

According to Smith (1995), the reason for cost estimating is to provide the most realistic prediction possible, of time and cost at any given stage in a project. The character and method of estimating changes as the project develops and the design approaches its final form. This happens because the unknowns can be shown to decrease as the project progresses until at the project completion the final cost is known with certainty (Smith, 1995). The first cost estimate that the client gets from the quantity surveyor is a very important factor to be considered in the client’s overall strategy of the decision to build. Smith (1995) suggests that the first estimate has a particularly crucial role to play because it is the basis for the release of funds for further studies and estimates. Even though the data available are limited at an early stage it is very important that the initial estimate is as accurate as possible. If the price given is very high this will discourage the client to move on with the project, while if the price is too low then the project will go out of the budget limits. Technology has played a significant role during the last decades in developing new methods for cost estimating which result to more accurate results, reducing at the same time the resources needed to reach to these results. One of the most widely used methods employing high technology, is cost modelling. Ferry and Brandon (1991) define cost modelling as a symbolic representation of a system, expressing the content of that system in terms of the factors, which influence its cost. The aim of cost models is generally to represent accurately the whole range of cost variables inherent in a building design to secure improved cost forecasts and design optimisation (Seeley 1996).

This paper reports on the results of continuing research implemented at UMIST to produce a model for the prediction of final building costs at an early stage in the project’s design phase.
PREVIOUS RESEARCH - THE PRODUCTION OF PROCOST

Ongoing research at UMIST commenced in 1997 has resulted in the production of ProCost, an early stage building cost estimating tool. The software is based on Artificial Neural Network technology to produce single figure estimates of the total building cost.

Research resulted in the identification of 41 variables which can affect the final price of a building project. These do not include only the traditional quantitative variables such as number of floors, type of construction etc, but also a series of attributes that would not normally be considered in a typical estimate but are shown to affect the final cost. Such variables are the contract type used, the tender strategy employed, or the procurement route followed. In general the input variables are separated into project strategic (such as procurement strategy, contract type), site related (such as topography, site access) and design related (such as substructure type, frame material etc).

Having identified all the variables, a series of data sets from past projects were collected from construction clients, quantity surveying and project management practices, as well as from BCIS subscribers. This resulted in the formation of a database of projects with all the cost-affecting attributes identified and with their final cost determined. The association between the input data (41 variables) and the output data (final building cost) was developed using Artificial Neural Network technology.

Artificial Neural Networks resemble the way the human brain works in order to be trained to learn how to give the appropriate output. The way that this is employed is that the network is presented with some input examples and their target outputs. What the network then does is to use the inputs and produce its own outputs. In the case of our concern for example, the neural network is presented with a series of building attributes at its input and the associated cost of the building at its output. If this is repeated for a number of projects then the network develops relationships between the different building attributes and cost, being able in effect to produce its own costs once presented with new variable values of a building to be built in the future. In total, the data collection programme resulted in the collection of 288 full data-sets of projects (Emsley et all, 2002). The final result of this research was the production of neural network based predictive model. In effect the user is able to input the value of each of the variables for the building under consideration in the program’s interface and to receive a prediction of its proposed cost. The output of the model is deterministic, that is the prediction is given in form of a single figure.

INVESTIGATING THE TRAGET MARKET REACTIONS

In 2002, a piece of research was initiated by the writers to investigate the reactions of potential ProCost users to the software and its use in practice. This was initially implemented in the form of a series of one to one semi-structured interviews with four representatives of the industrial collaborators of the research team (EC-Harris, BCIS, Symonds) who were using the software in conjunction with their own estimating methods. One of the primary areas for further development indicated as a common request amongst the different firms was the incorporation of an elemental cost estimation tool within the software. The opinion of the industrial collaborators has given an indication, however it is not enough to drive to conclusions relative to what the current trend of the Quantity Surveying profession is. Therefore it was considered critical to proceed to the next step of the research into conducting a nationwide questionnaire survey.

The questionnaire was sent in 120 QS practices in England, Scotland, Wales and the Northern Ireland. The response rate was 40%. It included a total of nine questions, all of which were in the form of a multiple-choice answer. The questionnaire was divided in two parts. Part one investigated the current elemental cost estimating practice, while part two was more specific to ProCost investigating ways in which the software could be improved. The results of the first part of the questionnaire have been published at a past paper (Soutos and Lowe, 2003). The second part was formed by three questions which are analysed extensively below.
As figure 1 denotes, more than half of the respondents declared that they would possibly use such software in the future. Such an answer descents from quantity surveyors’ general fear of technology. Postponing the use of high technology in estimating for the future implies that estimators know that it is advantageous and, therefore, do not deny it, but on the other hand, are very reluctant to make a change at once and incorporate high technology methods within their techniques. It is, in fact, another form of a negative answer as it indicates that the people who gave it would not currently use ProCost. Apart from them, a 19 percentile answered that they would not use a single figure price estimating software at all. This means that if ProCost was to be marketed currently, a total of 72% of the sample would not use it. This is a first indication that there is a need for a review of both the program interface and the general way on which it works.

Moving on to the second question, the results are indicated in the graph below.

With the answers on the previous question indicating the general need for change, a follow up question focuses on the change being initiated by the incorporation of an elemental breakdown tool into the software. 48% answered that they would “prefer” for the software to give an elemental breakdown rather than remaining in its current form. 51% replied that it is essential for this change to happen, demonstrating that they would not use the software unless the change was made. With half of the sample believing that the incorporation of an elemental breakdown tool would make it a more useful piece of software and the rest thinking that there is no practical point of using such software otherwise, it is clear that it is a general desire within the potential users group for this step to be implemented.
Would the inclusion of an elemental estimation tool encourage you to use ProCost?

Yes 94%
No 6%

Figure 3: Results on question 3

The last question has the form of a final confirmation and seeks to validate what was answered in the previous one. There are only two possible answers yes or no, directing the respondents to give clear positive or negative indications without the possibility of a middle answer. As figure 3 indicates, indeed a 94 percent of the population confirmed what was discussed above by answering that the elemental breakdown of the prices in the software would encourage them to use it.

The answers on these questions indicate that quantity surveyors cannot rely on single figure price estimating techniques. Incorporating therefore an elemental estimating tool within ProCost is vital in order to increase its target group by making it more useful and more in tune with the estimators’ current needs and practice.

METHODOLOGY OF ANALYSIS

Having verified the need for the incorporation of an elemental cost estimating tool within the model, the next step was to find out how this could be implemented. First of all a format for breaking down the building cost into different elements had to be chosen. A series of different techniques such as the Standard Form of Cost Analysis (SFCA) or the functional elements method, are used by different practices. Therefore an investigation into which particular method is the most widely used was vital in order to adjust the elemental format in the model according to the methods used in practice. This investigation was carried out by the first part of the questionnaire mentioned above. The results of this survey have been previously published (Soutos and Lowe, 2003), but in summary:

- 75% of the respondents use the BCIS SFCA as a method to produce their elemental estimates. The ones who do not use it are those that do not prepare elemental estimates often.
- 42% use the detailed level of analysis of the SFCA, while 38% use the top level at the initial stages of the estimate moving on to the detailed as the design evolves. Only 4% of the respondents use the detailed level only.

According to the results of the survey it is the detailed level of the SFCA that appears to be the most widely used way of elemental analysis of building costs. After careful consideration it was decided that implementing a model using this detail level would be very time consuming and difficult to employ as it is very hard to find fully developed analyses of finished projects. As a result the final format used for ProCost would be a hybrid level which can be produced from the detailed SFCA if fittings and external walls are taken out, as well as the sub-elements of services. The reason for pulling fittings and external works out of the model is because ProCost does not model for them. The services sub-elements were taken out for a number of reasons. The most important was that there can be great variations in them depending on the building type. In respect to the mechanical and electrical costs, the QS does not normally estimate the sub-elements but consults a services engineer for advice. The final format produced was as follows:
Having decided on a format to separate the building into elements, the next obstacle was to find a source that could provide a considerable number of past projects with the description of their elements’ properties and the provision of a cost for each one of them. The most obvious source was BCIS itself. The BCIS online system was used and 120 office development projects with fully developed elemental cost estimates where identified and inputted in a database. A long process of coding the attributes that ProCost’s Neural Network needs to define a building followed for the total of the 120 projects. Another database with the cost of each one of the elements was implemented. After the two database were formulated, they were inputted into SPSS for Windows. Regression analysis was used to produce a predictive model for each sub element. A total of 17 predictive models were therefore developed.

PRELIMINARY RESULTS

The results of the “superstructure” cost element are described as an example and some initial indications regarding the validity of regression analysis as a modelling technique for implementing the predictive tool are given.

The aim of the regression analysis for the superstructure element was to find the regression equation that would best describe the way that a series of project descriptive variables affect the cost of the superstructure of the building. Four types of regressions analyses were implemented, one forward regression, one backward, one forward excluding the envelope and wall to floor ratio and one backward excluding the envelope and the wall to floor ratio. The reason that these two variables were excluded on the last two analyses was the fact that out of the 120 projects in the database only a few had description of these two variables. The lack of these data could possibly affect the regression results.

After implementing the analyses all the models produced were studied carefully. The selected model was that with the greatest adjusted $R^2$. The model that best described the relationship of a series of descriptive variables with the cost value of the superstructure was finally identified to be model 17 of the backward regression including envelope and wall to floor ratio. This model produced an adjusted $R^2$ of 0.779. The results of this model are summarised in table 1.

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std. Error of the Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>.941</td>
<td>.885</td>
<td>.779</td>
<td>4.11891</td>
</tr>
</tbody>
</table>

The value of adjusted $R^2$ is significant as it means that 77.9% of the cases of the superstructure cost values can be explained by the regression equation. Having confirmed the
significance of the regression equation, this was used in order to predict the values of superstructure for the 120 projects of the database. The values obtained were then compared with the actual values of the superstructure cost for each one of the projects using SPSS and Excel for Windows. In comparing the predicted with the actual values, only 55 of the 120 projects analysed in SPSS were labelled valid. The reason for this is that even when one of the significant to the superstructure variables is missing from one project, SPSS cannot analyse it. The results of the comparison are shown in table 2 below.

Table 2: Superstructure: Comparison of actual and predicted values

<table>
<thead>
<tr>
<th></th>
<th>Predicted Values</th>
<th>Actual Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>N Valid</td>
<td>55</td>
<td>120</td>
</tr>
<tr>
<td>Missing</td>
<td>65</td>
<td>0</td>
</tr>
<tr>
<td>Mean</td>
<td>40.0542</td>
<td>42.0119</td>
</tr>
<tr>
<td>Median</td>
<td>39.5397</td>
<td>41.1395</td>
</tr>
<tr>
<td>Mode</td>
<td>16.88</td>
<td>16.45</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>11.11167</td>
<td>9.27271</td>
</tr>
<tr>
<td>Minimum</td>
<td>16.88</td>
<td>16.45</td>
</tr>
<tr>
<td>Maximum</td>
<td>64.47</td>
<td>92.27</td>
</tr>
<tr>
<td>MAPE</td>
<td></td>
<td>9.2</td>
</tr>
</tbody>
</table>

The values in this table refer to the percentage of the total building costs that correspond to the superstructure. For example the superstructure actual costs are on average 42.01% of the total building cost. As it can be seen from this table, the predicted values of the superstructure are very close to the actual ones. It is worth noting that the mean of the predicted values is 40.05, while the mean of the actual is 42.01, giving a variance of the means of only 4.7%. The value of MAPE is the mean absolute percentage error, which is produced by finding the absolute values of the difference between the actual and the predicted values and therefore the absolute percentage error for each project and then finding the mean of all these errors.

In order to analyse the variance between the predicted and actual values of individual projects, the scatter gram shown below is produced.

Figure 4: Scatter gram showing the percentage error between the actual and the predicted value for each one of the projects tested
This graph indicates the exact percentage of variation of the predicted value from the actual. It can be observed that the variation for individual projects can be from almost zero to about 50%. The amount of projects having a certain variable range is illustrated in figure 5.

![Percentage of projects having a certain variance range](image)

**Figure 5: Percentage of projects having a certain variance range**

According to this more than a quarter of the projects have been predicted very accurately varying only between 0 and 5% of the actual cost, while about 55% of the projects were predicted with a variance of 0-10% of the actual cost. Of the 56 projects only 7 were predicted with a cost values bigger than 25% of the actual value. These results look very promising and give an initial indication that a robust model for predicting the elemental cost of a building can be produced.

**CONCLUSIONS**

Table 3 indicates schematically the analysis procedure and the development steps followed up to this point in the research.

<table>
<thead>
<tr>
<th>Research Step</th>
<th>Methodology</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Investigation of the reactions of the target market to identify needs and potential for improvement.</td>
<td>Nationwide questionnaire survey</td>
</tr>
<tr>
<td>2</td>
<td>Investigation of the best format to be utilised for the breakdown of elemental costs.</td>
<td>Nationwide questionnaire survey</td>
</tr>
<tr>
<td>3</td>
<td>Formulate two databases from 120 office development projects. One with the projects' attributes, the other with the costs of all the elements of each project.</td>
<td>BCIS online system</td>
</tr>
<tr>
<td>4</td>
<td>Implementation of the predictive model (Regression analysis)</td>
<td></td>
</tr>
</tbody>
</table>
The prediction of elemental building costs during the initial stages of the estimates is very important to the work of the estimator. This has been identified through a series of interviews and questionnaire surveys.

Ongoing research in the Project Management Division of the Centre for Civil and Construction Engineering at UMIST has resulted in the production of ProCost, a Neural Network based software which has the ability to predict in the form of a single figure, the cost of a building at a very early stage in the design process. The fact that the software cannot predict elemental building costs is identified as a major drawback which has to be eliminated in order for the program to apply to wider range of the market and meet the needs of the majority of the estimators.

As a result it was decided to go on with the second stage of the research at UMIST which would have as an objective the incorporation of an elemental estimating tool within ProCost. After a nationwide analysis of current cost estimation practices the most widely used way of separating buildings into elements was identified and used as the basis for producing an elemental estimating tool.

Regression analysis methods were utilised in order to produce a predictive model for each one of the seventeen elements identified.

The results of the regression analysis carried out to indicate the cost of the superstructure, were presented. The initial indications of the regression analysis seem to be very promising. The results of the predicted values are very close to the actual ones with a variance of the mean of all the projects at only 4.7%. The mean absolute percentage error (MAPE) of all the projects analysed is 9.2%. This suggests that an accurate regression based model can be developed to predict the costs of each of the elements described above. Integrating such a tool within the ProCost software is definitely a great challenge. Nevertheless, if this could be put into practice the result would be a powerful elemental cost estimating tool which could prove to be an invaluable tool for the early stage design cost estimator.

REFERENCES


Soutos M., Lowe D. (2003), An investigation into the current practice of elemental cost planning, Proceedings The1st ProBE conference, Glasgow Caledonian University, UK