THE COST OF PROCUREMENT: A NEURAL NETWORK APPROACH

Anthony Harding, David Lowe, Adam Hickson, Margaret Emsley and Roy Duff
Department of Building Engineering, UMIST

ABSTRACT: Existing research that has attempted to determine differences between the costs of following different procurement routes has consistently aimed to determine a single blanket figure, such as "design and build is 15% cheaper than traditional". No attempt has been made to provide a difference which is project specific (Duff et al., 1998). Furthermore, no previous research has determined the total cost to the client using any objective method.

The lack of data defining client costs and the absence of suitable modelling techniques have prevented such an objective evaluation being made (Masterman, 1994). These factors prompted research into the cost of procurement at UMIST. This research has required the collection of a substantial database of the total cost, to a client, of past projects, and the subsequent creation of a neural network model of these costs.

Results of the first phase of development of this model are presented, including regression analysis, preliminary neural network models and sensitivity analysis. An assessment of how these results will inform future development of the model is also made.

KEYWORDS: Cost Modelling, Early Stage Estimating, Neural Networks, Procurement.

1. INTRODUCTION

The selection of the most appropriate procurement route is believed to have a significant influence on the final cost of a building project. Despite this, however, very little research has aimed to provide any indication of the relative costs of this important strategic decision. Research that has been carried out has tended to be quite subjective, and often fails to consider the total cost to the client. In addition, previous research has aimed to provide a single figure which represents the proportional difference between one procurement route and another. This assumes that the differences in are constant for all projects, an assumption which has not been validated in the literature.

This paper reports preliminary results of the analysis phase of a research project which aims to provide a model of the relative costs of procurement routes. The final model will be able to determine the total cost, to the client, of a building project, taking into account all cost significant variables which are known at the early stage of the project. As procurement is one of these significant variables, this model would be capable of predicting the costs of the project under different procurement routes. Thus it can provide the cost of following different procurement routes for any project, and thereby inform the client of the cost of following different procurement routes.

In this early phase of analysis, regression and neural network analysis of a subset of the data is presented. While the final model will determine the total cost – to the client – of the project, in this paper only the final account contract cost is modelled, as the set of client's costs is still being collected. Nevertheless, the analysis does provide a useful perspective on development of the model, as well as bringing a number of important modelling issues to the fore, which have implications for early stage cost estimation in general.

2. PREVIOUS RESEARCH

The costs associated with procurement were first evaluated subjectively. The Department of Industry and the Department of the Environment (Department of Industry & Department of Enviroment, 1982) made a subjective comparison between traditional, design & build and management procurement routes. Differences in cost were not quantified but subjectively evaluated as "higher" or "lower" than the traditional procurement route.

A subjective comparison was also made by Brandon et al. (1988), who suggested that the cost of the procurement route be taken into account by means of an addition. The additions range from 0% for the conventional and design & build systems to 15% for construction management, apparently based on experience.

The first attempt at an objective evaluation was made in a recent report on design & build (Reading University, 1996). The Centre for Strategic Studies in Construction concluded that the design & build procurement system is "at least 13% cheaper than traditional procurement". The analyses were carried out using multiple regression and identified 11 variables, including choice of procurement system, but which together explained only 51% of the variability in project cost.

A more recent evaluation of the cost of the procurement route was made by Elhag et al. (1998). They compared the average tender prices per m^2 for the three different procurement systems for 13 office projects and 28 industrial projects. The price per m^2 in the office buildings was found to be highest for management contracts, and lowest for the traditional route. For industrial buildings, design and build was found to be cheaper than traditional, the reverse of what was found for the office projects. This suggests that the relative costs of different procurement routes are not constant for all projects.

Unfortunatley, the results of this analysis cannot be relied upon with much confidence. No test of statistical significance was performed on the results, the number of projects is very small and the comparison did not take the possible influence of other cost significant variables into account. Nevertheless, it shows that the assumption that one procurement route is a fixed percentage cheaper or more expensive than another may be flawed.

3. MODEL

In order to have a valid model of the cost of following different procurement routes, it is necessary to have a model which will not assume that these differences in cost are fixed, but that they are a function of the other cost significant variables. The relative cost of different procurement routes must be determined specifically for each project. The simplest way of determining this cost difference between procurement routes is to create a model that is able to predict the cost of the project from all the cost significant variables, including procurement. This means that not only is the direct effect of procurement on cost modelled, but also how that effect varies as other cost significant variables vary.

This model can then be used to accurately evaluate the expected costs of the project under different procurement routes, from which the differences between the predicted costs can then be obtained.

In order to accurately evaluate the differences in cost between the different procurement routes for any particular project, it is necessary to model the complex and little understood interrelationships which exist between the cost and all the cost significant variables that have been identified as possibly significant at this early stage of the project. Because the relationships between the different variables are so difficult to identify and quantify, a neural network has been selected as the modelling tool to use. Neural networks are able to model any function to an arbitrary accuracy without those functions having to be identified in advance, and are therefore ideal for use when the interrelationships between variables are not understood. They have the ability to implicitly quantify all the interrelationships between the variables by examining the outcomes of a large set of past projects. This process is called "training".

Many of the modelling issues regarding the application of neural network modelling of this problem have already been discussed in some detail in a previous paper from the project (Harding et al., 1999; Harding et al., 2000). These show how the model is required to predict the cost of the building only, rather than the whole cost of the building. This means that items which are additional to the building , namely the fittings and external works, must be removed from the cost of any building project collected. In addition it is also necessary to remove the influences of time and location using the BCIS cost indices to bring the projects to a common location and base date. Thus, when predicting the cost of a building, the model predicts the cost for the standard time and location. This cost can then be adjusted, using BCIS indices, for the time and location in question.

4. ANALYSIS

The collection of data from past projects, which includes cost information and the forty potentially cost significant variables already identified (Harding et al., 1999), is nearing completion. Since identifying the forty variables, two variables (sanitary installation and disposal installation) are no longer considered because almost no variation in their definition was found among the projects collected.

4.1 Description of the data set

The data set used for the preliminary analysis comprises 58 projects for a variety of building functions (administrative, residential, industrial etc.). The costs of these projects varies between £44.24K and £4.64M, with gross internal floor areas between 154m^2 and 11605m^2. Of the 58 projects, 34 are design and build projects, and the remaining 24 were procured by the traditional route.

While the model aims to predict the cost of the building, representing this cost to the neural network has presented some problems. The cost of building projects varies significantly. This is apparent in the data set, where the largest, most expensive project is over 100 times the cost of the smallest. Because neural networks seek to minimise the absolute error in the prediction, a large error in an expensive building will be much more significant than a large error on a smaller building. The usual solution to this problem is to use the log of the value, rather than the value itself. This also tends to avoid the problem of all the values being grouped together at one end of the scale.

However, there is another way to represent cost which might be used: the cost per square metre. The most common practice in early stage cost estimation is to determine the most likely cost per square metre of the building, usually using some kind of interpolation between similar buildings whose cost is already known. Using the cost per square metre would mean that the linear effect of the gross internal floor area (GIFA) on the project cost – an effect which is already understood and accounted for – would be effectively removed from the model. This allows any further modelling to be focussed on other relationships, which are perhaps not so well understood. Focussing on these weaker, but still significant relationships could be particularly effective for the neural network model, which might have difficulty in finding these weaker relationships in the presence of a relationship which is as strong as that between cost and GIFA. It should also be noted that while the strong, linear relationship between cost and GIFA is removed, the neural network would still be able to model any non-linear aspects of the relationship between cost and GIFA which might exist, provided GIFA is included as an input relationship.

4.2 Regression

Before attempting a neural network analysis it is very useful to try to identify what linear relationships can be identified from the data set. Therefore a simple regression analysis was performed. This analysis involved the creation of three regression models, which predicted the cost, the log of the cost and the cost/m^2 respectively. In order to create these models, a stepwise selection method was used, such that a variable could be added if its value of t was significant at 95% confidence.

The three models created were an 11 variable model for cost, a 9 variable model for the log of the cost, and a two variable model for the cost/m^2. In order to increase the number of cost significant variables considered by the cost/m^2 model, the confidence level was dropped to 90%. This yielded a 10 variable model.

None of the three models determined included procurement, so the procurement route was added to each model to assess its potential impact on the cost. The models are shown in Table 1, along with the increase in R^2 which was observed on the addition of each variable.

Cost		Log of cost		Cost/m²	
Variable	Increase in R^2	Variable	Increase in R^2	Variable	Increase in R^2
GIFA	0.833	LN(GIFA)	0.899	Function	0.233
Storeys Above	0.059	Internal Doors	0.024	Internal Doors	0.145
Site Nature	0.015	Function	0.020	Storeys Above	0.056
Shape	0.013	Storeys Above	0.008	Units	0.055
External Walls	0.010	LN(Storeys Above)	0.007	Piling	0.053
Internal Doors	0.009	Windows	0.005	Substructure	0.042
External Doors	0.008	Location	0.004	Stairs	0.037
Height	0.007	Height	0.004	Topography	0.032
Function	0.005	Topography	0.003	Ceiling Finishes	0.028
Stairs	0.005	Procurement	0.000	Frame	0.021
Substructure	0.004			Procurement	0.002
Procurement	0.000				

Table 1. Variables in the regression analyses

The results from the cost and log cost models show that the GIFA is by far the most significant influence on cost. The influences of other variables are not nearly as significant. This was expected. However, what is perhaps surprising is that the cost significant variables differ from model to model. Some variables are common to all three models (function, storeys above and internal doors), and some appear in two of the models (height, stairs, substructure and topography). The remaining 11 cost significant variables which appear are unique to one model.

In order to assess the implications of this finding, it is necessary to consider the nature of the model. The three outputs for the model differ in nature, and that influences how the variables influence the final cost. If a variable causes an increase of one unit in the raw cost model, then the raw cost rises by that amount. This means that variables which influence this model should, intuitively, be those which contribute linearly to the cost. In the log cost model, on the other hand, if a variable causes in increase of one unit within the model, the net effect on the actual cost is a proportionate one. The cost/m^2 model is somewhere between the two, as if a variable causes an increase in one unit within the model, then the corresponding increase is multiplied by the GIFA to obtain the influence on the total cost. As the GIFA is approximately proportional to the cost, this means that the influence of the variable is approximately proportional to the cost.

Thus one would expect that the variables found by the different models would be different. The variables found by the raw cost model would be those whose influence on the raw cost has a linear component, and those variables found by the log cost model would be those whose influence on the raw cost has a strong proportional component. The cost/m^2 model, on the other hand, shows which variables strongly influence cost once the strong linear relationship between cost and GIFA has been removed. Nevertheless, it does have variables in common with both of the other two models, which is consistent with its position as a model whose influences are neither wholly proportional nor wholly linear.

While it might be reasonable to accept the variables from the log cost and cost/m^2 models, it has been asserted that those variables which are used in the raw cost models are those which have an influence on the raw cost. However, apart from the height of the building and the number of storeys and the height, it is difficult to see how any of the other variables should affect the model in some way. It would appear likely, then, that the reason some of these variables appear, is because they tend to correlate well with the building cost – that is to say, higher (or lower) values of that variable tend to be associated with higher costs. This does not necessarily mean that these variables do drive the cost of the project in a direct (rather than a proportional) manner.

A second important thing to note about the results of the analysis is that the apparent increases in R^2 achieved by adding variables to the model are not necessarily measures of their influence on the cost. Regression analysis assumes that all the variables presented to it are completely independent. However, this is not necessarily the case. Indeed, some of the variables were quite well correlated. For example, internal doors is one of the most significant cost indicators for all three models, yet it would not usually be considered to be a significant cost driver. However, internal doors correlates quite well with all the other internal finishes, yeilding correlation coefficients of between 0.25 and 0.5 with the other finishes. Thus it can be seen that internal doors could actually be considered as a fairly good measure of the overall level of the internal finishes, a variable which *is* cost significant. This means that some variables may appear to be more cost significant than they actually are. Conversely, it also means that the cost significance of other variables might be underestimated, as some of their significance in the model might already have been implicitly included in the model through other variables with which the variable is correlated.

As well as addressing the type and significance of variables added, it is also important to consider the relative performances of the three models. These are shown in Table 2. From the R^2 values of the respective models, it would appear that the cost/m^2 model is not as good an indicator of cost as the other two models. However, when the model's predictions are expressed as a raw cost (i.e. its predictive capabilities are expressed in the same terms as the raw cost model, which is indicated by *R^2 on cost* in the table), it becomes clear that the cost/m^2 model is actually a better predictor of cost than the raw cost model.

Model		R^2	R^2 on cost	MAPE	Median APE
Without Procurement	Cost	0.968	0.968	67.3%	48.2%
	Ln(Cost)	0.974	0.985	13.8%	12.1%
	Cost/m^2	0.702	0.979	15.6%	11.0%
With Procurement	Cost	0.968	0.968	64.9%	44.7%
	Ln(Cost)	0.974	0.985	13.9%	12.2%
	Cost/m^2	0.704	0.979	15.4%	10.1%

Table 2. Performance of regression models.

In addition to the R^2 values, two other performance measures are provided: Mean Absolute Percentage Error (*MAPE*) and Median Absolute Percentage Error (*Median APE*). These show the mean and median error (expressed as a proportion of the cost) respectively. They show how the high R^2 value indicated for the raw cost model is misleading, as while the cost of the high cost projects is accurate this accuracy does not extend to the low cost projects. The reasons for this were discussed in section 4.1. These two average percentage error values will also be useful in comparing the performance of the neural networks to the regression models.

The implications of the results of these models are that procurement does not have a significant linear influence on the cost of building projects. The introduction of procurement to the models only improves the cost/m^2 model, and even then not significantly (the value of t for procurement is 0.559).

4.3 Neural network analysis

If there is no apparent linear or proportional relationship between the procurement route and the cost of the building then, if the procurement route does significantly influence the cost, the relationship must be predominately non-linear. In order to assess this, neural network analysis was performed. Models were trained to predict two measures of cost: the log of the cost and the cost/m^2. The raw cost was not modelled for reasons already outlined.

The data set was split into a training set of 32 projects, a verification set of 13 projects (to prevent overtraining) and a test set of 13 projects (for independent validation of the model). A wide variety of possible input variables and architectures were evaluated, and the best selected in two categories: those which included procurement as an input, and those which did not. This yielded four networks, two for each cost predictor. Their relative performances on the test set are shown in Table 3, along with the network architectures and inputs.

		R^2			Testing		
Model		Training	Verification	Testing	**R^2 on cost**	**MAPE**	**Median APE**
Without Procurement	Ln(Cost) [a]	0.971	0.992	0.837	0.933	25%	18%
	Cost/m^2 [b]	0.718	0.865	0.348	0.538	31%	22%
With Procurement	Ln(Cost) [c]	0.969	0.990	0.843	0.945	23%	13%
	Cost/m^2 [d]	0.702	0.859	0.252	0.696	27%	25%

[a] 3 layer MLP (5-4-1). Inputs: location, Ln(GIFA), storeys above, height, internal doors
[b] 3 layer MLP (10-3-1). Inputs: function, piling, internal doors, ceiling.
[c] 3 layer MLP (6-7-1). Inputs: procurement, location, Ln(GIFA), storeys above, height, internal doors
[d] 3 layer MLP (11-3-1). Inputs: Inputs: function, procurement, piling, internal doors, ceiling.

Table 3. Performance of neural networks.

The values of R^2 show that the neural networks are not generalising well. This can be seen from the fact that the R^2 value of the test set is much less than the verification and training sets. The accuracy of the neural networks on the independent test set can also be seen to be low from the value of MAPE. These values are lower than those obtained in the regression analysis. In addition to the values of R^2 and MAPE, the poor generalisation can also be observed from the fact the best networks for predicting log cost use different variables as input to those which predict cost/m^2. This shows that the networks are merely identifying some of the easily quantifiable trends in the data rather than modelling the real relationships in terms of the variables which are truly the most cost significant. In order to increase this generalisation performance, it will be necessary to use a much larger data set. Only then might the model be used as an indicator of project cost.

Despite the fact that the model is not generalising well, the models which contain procurement do appear to yield a slight improvement in the model both in terms of the *R^2 on cost* and *MAPE*. However, the results of the R^2 and *Median APE* do not consistently show those networks which included procurement to be better. Nevertheless, given that increasing the complexity of a network tends to decrease its accuracy, the fact that the network shows a slight improvement, rather than a slight deterioration, when procurement (an additional input) is added, suggests that the procurement does have some effect on the accuracy of the model. This suggests that there is some relationship between procurement route and cost, although it is difficult to say with any certainty how significant this relationship might be without using more data.

4.4 Implications for procurement route

While the low values of R^2 show that the preliminary model is not accurate enough to be used as a cost predictor, it would be useful to know how the neural network is taking the procurement route into account, and whether it is consistent between the two networks. If the two networks that included procurement were modelling a strong non-linear relationship between the cost and the procurement route, then both networks would be expected to have "learned" that relationship. If this is the case, then changing the procurement route for any project should yield similar results for both of these networks.

In order to assess this, the two networks were presented with two new data sets. These data sets were essentially the full data set, except that all the procurement routes were all set to traditional for one set, and set to design and build for the other. For each project, therefore, the predicted increase in cost associated with following design and build, rather than traditional, was determined and compared. These increases were expressed as a proportion of the predicted cost of the traditionally procured project, and are summarised in Table 4, along with the correlation.

	Mean Increase	Mean Difference
Ln(Cost)	0.2%	5.5%
Cost/m²	-4.5%	10.3%

Correlation: -0.08863

Table 4. Increases in project cost between traditional and design and build.

The results of the analysis show quite clearly that both networks appear to be modelling some kind of relationship between procurement route and cost. The fact that the mean difference is greater than the mean increase (showing that some increases are negative, i.e. design and build is cheaper than traditional) also shows that this relationship is not constant from project to project. However, if the networks were modelling the same, real relationship, then the increases would show a strong positive correlation. The correlation between the two is, however, a weak negative one, suggesting that the networks have not found a real relationship between procurement route and cost. Thus, at this stage it can be seen that poor generalisation prevents the assessment of any relationship between procurement route and cost.

5. FUTURE DEVELOPMENT

The results of this preliminary analysis, while producing useful indicators of some of the modelling issues which might be faced in the analysis of the completed data set, used only a small data set. The completed data set is expected to comprise between 400 and 500 projects. The analysis and modelling of this data set will expand considerably from the analysis presented here. Some of the principal areas in which the analysis will be expanded are as follows.

5.1 Client costs

In order to model the whole cost of the project to the client, it is also necessary for the model to include the client costs. Two approaches to this will be analysed: modelling the client costs using a second neural network (using a similar approach to that used for the contract cost), and modelling the whole cost as a single figure using a single neural network. While it might appear from the preliminary analysis that there is little relationship between the contract cost and the procurement route, there may be a stronger relationship between client costs and procurement. Intuitively, this might be expected, if only because the client will have reduced design fees in a design and build project.

5.2 Cost per m^2 and total cost

The preliminary analysis suggests that the model may best be modelled using the log of the cost. However, the cost/m^2 performed similarly, and one may prove more suited to analysis on a larger data set than the other. Therefore, modelling and analysis should be continued with both.

5.3 Data reduction techniques

A simple regression analysis can be useful in determining which variables might be most significant in determining cost, and what linear relationships can be identified. However, it may be possible to use both factor analysis and cluster analysis in order to help identify where interrelationships between the defining variables lie. This would be particularly useful when trying to determine which relationships between a variable and the cost are real relationships, and which are merely apparent relationships, arising because certain values of a variable trends to be associated with particular types of project. Additionally, they could help identify which variables are cost significant, and which are not. These results could aid the modelling process, as well as providing an insight into projects which could be useful in itself.

6. APPLICATION OF THE MODEL

While the driver for the research is the comparison of the costs of different procurement routes, it involves the creation of a robust cost estimation tool. If the accuracy of this tool were sufficient, then it could prove an important addition to the early stage cost estimator's toolbox, providing accurate, objective estimates of the likely cost of future projects.

In addition to this, the completed model – if successful – could allow the comparison of the cost of different procurement routes. However, even if no relationship between the procurement route and the cost can be identified, there may be differences in cost associated with other strategic variables, such as the tender strategy and planned duration. These costs would constitute useful strategic information for the client. Additionally, such information would be useful from a research point of view to see how, typically, the changing of these variables affects the cost of different types of project.

This type of comparison is not restricted to those variables that represent strategic choices for the client to make. It can also be useful for the comparison of different types of project function, and also for comparing speculative and owner-occupied buildings. Indeed this comparison could expand beyond project strategic variables to issues of location, site, type of frame, et cetera. Comparison of such options would permit the identification of how and to what extent each of these variables influences the project costs.

In addition to information that might be obtained from the model, other information is also being collected whose analysis might prove useful. The first is the tender price. By modelling this it would be possible to compare the differences in tender price and final account and how they vary from project to project. Similarly, the planned project duration could be compared to the actual duration, which would provide an indication of whether certain types of project are more likely to overrun than others.

By comparing these two values it may be possible to quantify both cost and duration certainty. This would permit the client's advisors to provide objective advice on the cost risk associated with a particular project.

7. CONCLUSIONS

The selection of procurement route is believed to significantly influence the cost of a building project. However, very little analysis has been done to assess this, and much of this analysis has tended to assume that differences in cost between different procurement routes can be expressed as a fixed percentage of the cost of the project, and that this applies to any building project. This assumption has never been verified.

In order to test this assumption, it has been proposed to construct a neural network model of the total cost of construction, to the client. By including procurement as one of the cost significant variables, it would then be possible to assess the relationship between the procurement route and the cost of the project.

Preliminary analysis of the final account cost of 58 building projects was performed. This analysis attempted to verify the assertion that the procurement route has a significant influence on the cost. Firstly, it was attempted to find, using multilinear regression, a linear relationship which showed that design and build was a fixed amount or fixed percentage cheaper than the traditional procurement route, or vice-versa. Such a relationship was not found. Only a very weak relationship was identified in the cost/m^2 model, and it was found not to be significant.

Accepting that no linear relationship could be identified, it was attempted to identify a non-linear relationship using neural network analysis. Results of this analysis did suggest that the inclusion of the procurement route within the model might make it more accurate. However, the models did not generalise well. An investigation of the relationships between procurement and cost that the networks appeared to be learning revealed that there was no consistency between the two networks, suggesting that this poor generalisation was preventing the networks form effectively learning any relationships which might exist. However, it was explained that once the completed data set was modelled, the generalisation would be significantly better, allowing the relationship between cost and procurement route to be modelled much more effectively.

The analysis presented is only preliminary, and has been performed only on a small subset of the full data set which will be analysed. It is anticipated that the analysis of the full data set will expand upon that presented here. Therefore some of the areas in which the analysis might be expanded were discussed.

Finally, three potential applications of the final, working model were presented:

- An objective cost prediction tool, which would be useful in generating accurate early stage estimates.
- A cost advisory tool, which is able to show the difference between different important project options.
- A research tool, where the model is used to identify how changing different options affects the costs of different types of project.

8. ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of the EPSRC and Mike Turega – for his help on some of the more difficult aspects of neural network modelling.

REFERENCES

Department of Industry & Department of Enviroment (1982), The United Kingdom Construction Industry - a guide to methods of obtaining a new industrial building in the UK, Invest in Britain Bureau, London.

Duff, R., Emsley, M., Gregory, M., Lowe, D., Masterman, J. (1998), Development of a model of total building procurement costs for construction clients, 14th annual conference, ARCOM, Reading, 210-218.

Elhag, T. M. S., Boussabaine, A. H.. (1998) Statistical analysis and cost models development, EPSRC Research Grant Report, University of Liverpool.

Harding, A. M., Lowe, D. J., Hickson, A., Emsley, M. W., Duff, A. R. (1999) Implementation of a neural network model for the comparison of the cost of different procurement approaches, 15th Annual ARCOM Conference, Liverpool John Moores University, 763-771.

Harding, A. M., Lowe, D. J., Hickson, A., Emsley, M. W., Duff, A. R. (2000) Implementation of a neural network model for the comparison of the cost of different procurement approaches, CIB W92 Construction Procurement System Symposium, Santiago, Chile, 24-27 April 2000, pp. 269-280.

Masterman, J.W.E. (1994), A study of the bases upon which clients of the construction industry choose their building procurement systems, unpublished PhD thesis, UMIST, Manchester.

Reading University (1991), Construction Management Forum, Report & Guidance, Centre for Strategic Studies in Construction, Reading.