Construction Management and Economics (2002) **20**, 465–472

Data modelling and the application of a neural network approach to the prediction of total construction costs

MARGARET W. EMSLEY*, DAVID J. LOWE, A. ROY DUFF, ANTHONY HARDING and ADAM HICKSON

Manchester Centre for Civil and Construction Engineering, UMIST, PO Box 88, Manchester M60 1QD, UK

Received 17 January 2001; accepted 19 April 2002

Neural network cost models have been developed using data collected from nearly 300 building projects. Data were collected from predominantly primary sources using real-life data contained in project files, with some data obtained from the Building Cost Information Service, supplemented with further information, and some from a questionnaire distributed nationwide. The data collected included final account sums and, so that the model could evaluate the total cost to the client, clients' external and internal costs, in addition to construction costs. Models based on linear regression techniques have been used as a benchmark for evaluation of the neural network models. The results showed that the major benefit of the neural network approach was the ability of neural networks to model the nonlinearity in the data. The 'best' model obtained so far gives a mean absolute percentage error (MAPE) of 16.6%, which includes a percentage (unknown) for client changes. This compares favourably with traditional estimating where values of MAPE between 20.8% and 27.9% have been reported. However, it is anticipated that further analyses will result in the development of even more reliable models.

Keywords: Cost modelling, neural networks, linear regression analysis

Introduction

The importance of models to estimate the cost of buildings has been highlighted by Ferry *et al.* (1999). Newton (1991) reviewed over 60 cost models and classified the techniques used to develop each model under eight headings, including regression techniques. However, in both cases no mention was made of the application of neural networks. Elhag and Boussabaine (1998) developed neural network models to predict the tender price of school buildings, and later (Elhag and Boussabaine, 1999a) they developed two models to predict the tender price of office buildings using linear regression and neural network techniques. They found that both techniques produced models that were able to map the underlying relationship between the cost factors and the tender price but, because the sample size was small (30 and 36 projects, respectively), concluded that more projects were required for meaningful conclusions to be drawn.

This paper describes the development of neural network models of total construction project cost based on recent historical project data. The initial impetus for the research was the paucity of data available that can provide reliable information about the relative costs of using different procurement routes. However, in attempting to develop a model to address this strategic decision, it immediately became apparent that this variable cannot be isolated from the many other cost significant variables in a building project (Harding

*Author for correspondence. e-mail: margaret.emsley@umist.ac.uk

Construction Management and Economics
ISSN 0144–6193 print/ISSN 1466-433X online
http://www.tandf.co.uk/journals
DOI: 10.1080/01446190210151050

et al., 1999a), and a model is required that incorporates all the cost significant variables, the values of which are known at the early stage of the project.

This work has been carried out in two stages.

- An initial pilot study was made, where potentially cost significant variables were identified, the availability of data was investigated and strategies for data collection were established. In addition appropriate modelling strategies were examined, and preliminary testing of these methods was carried out, using a relatively small number (46) of data sets (Duff *et al.*, 1998).
- A full scale study was made using data from nearly 300 projects, and hence addressing one of the deficiencies in the model developed by Elhag and Boussabaine (1999a), in which more sophisticated models were developed.

Both the data requirements and the data collection processes are described, and ways in which both the input variables (such as frame type) and output variables (cost) may be best represented in the model are discussed. For the purposes of comparison, linear regression models have also been developed, and the results obtained are given before the development of the neural network modelling process is explained. The first two sets of models to be developed use first the five and then the nine most significant variables identified by the linear regression modelling process, and the final set incorporates all the variables.

Data requirements

The model variables may be divided into input variables and output variables. Initially, 43 input variables were identified, subsequently reduced to 41, as two variables were eliminated (sanitary installations and disposal installations) because almost no variation in their definition was found among the project data collected. Input variables were further categorized as project strategic, site related or design related (Table 1).

The identification of potentially cost significant variables was achieved through a thorough literature search of over 60 publications, supplemented by discussions with the professional collaborators (see Acknowledgements). In addition, with the exception of 'quality of building', all other input variables are encapsulated in the cost analyses published by the Building Cost Information Service (BCIS). This is less than the 67 variables identified by Elhag and Boussabaine (1999b), but it should be noted that that list also includes factors affecting construction project duration and factors classified as contractor attributes, which would not be known at the stage when the present model is intended to be used.

A criticism of previous cost models is that they used only the tender price to evaluate cost, whereas in reality the cost to the client of a building contract is the final contract sum. This is very rarely the same as the tender price, and Corbett and Rowley (1999) suggested that the final account sum should be made available to cost planners, whereas the BCIS, for example, provide only the tender price. The model described here has been developed using final account figures as the output variable. In addition, the whole cost to the client includes not only the final contract sum but also the professional fees and whatever resources the client has provided to the project. Models were developed separately for construction costs and client costs, but they may be summed to give the total project cost to the client.

The variables of time and geographic location were accommodated through the application of the BCIS cost indices to bring all projects to a common location and base date. The costs predicted, using the model, were then adjusted by the appropriate indices for the time and location in question.

Where projects included external works, demolition, fittings or specialist services, their associated costs were removed from the final account figure and the appropriate proportions were removed from the contract preliminaries and clients' costs (Harding *et al.*, 2000a). This was done because these costs are subject to wide variation, largely independent of the main variables defining the building. For example, for the projects where data were collected, the cost of external works varied from 1% to 30% of the total contract sum. Such variation makes these features impossible to model accurately and they are more reliably estimated independently.

Data collection

In total, the data collection programme resulted in the collection of 288 full data sets from predominantly primary sources, supplemented by some secondary data.

Primary sources

The professional collaborators provided a great deal of the data required, and contact was established also with organizations, primarily quantity surveying and project management practices, that were willing to provide data. A data pro-forma was developed to assist both the researchers and, more importantly, those collaborators willing to carry out the data retrieval themselves.

Table 1 Classification of input variables

Project strategic variables:		
Contract form	Procurement strategy	Quality of building
Duration	Purpose	Tendering strategy
Site related variables:		
Site access	Type of location	
Topography	Type of site	
Design related variables:		
Air conditioning	Internal doors	Roof profile
Ceiling finishes	Internal walls	Shape complexity
Electrical installations	Internal wall finishes	Special installations
Envelope	No. lifts	Stair types
External doors	No. storeys above ground	Substructure
External walls	No. storeys below ground	Structural units
Floor finishes	Mechanical installations	Upper floors
Frame type	Piling	Wall-to-floor ratio
Function	Protective installations	Windows
GIFA	Roof construction	
Height	Roof finishes	

This method of collection provided the great majority of building cost analyses. Thirty-nine offices were visited from 20 different organizations across the United Kingdom.

Secondary sources

The BCIS publishes cost analyses for construction projects, and these fulfilled the data requirements, except for the:

- final account;
- actual duration;
- quality of building;
- clients' external costs; and
- clients' internal costs.

In order to obtain this additional information, a questionnaire was sent to BCIS subscribers, yielding 29 sets of data.

In addition to these questionnaires, data were obtained from a much more extensive mail-shot administered to 1239 practising quantity surveyors, all of whom had been canvassed by telephone, but this yielded only six additional projects.

Data representation

Each of the variables has been analysed in order to determine the best way of representing that variable in the modelling process. The ways these variables are represented fell into four distinct groups.

The first of these groups comprised variables that are real numbers, for example, 'duration' and 'no. lifts'. Where the range of these variables differed by more than one order of magnitude it was more appropriate to use the logarithm of that value, to ensure that the range of values was more evenly distributed.

The remaining variables are categorical variables that represent one of a choice of categories. As a general rule it is best that a single input is used for a variable only when that variable has some meaning as a single variable (Tarassenko, 1998), i.e. if the value of the variable increases then it must represent an increase in some factor that influences the outcome of the model. For some variables, obtaining such an order was simple. For example, clearly with 'site access' there is an order between 'unrestricted', 'restricted' and 'highly restricted', inasmuch as an increase in the restriction to access will be expected to cause an increase in cost. Therefore this variable can be represented by a single input.

There were a great many more variables for which no such order was immediately apparent, e.g. 'internal wall finishes', where the variable represents the cost of different material combinations that will make up the finish. The value of the input was set to be the standard cost per m^2 of each finish, which provides an order proportional to how much each finish is expected to impinge upon the final building cost.

For some categorical variables a consistent order could not be identified, because the actual differences in cost between the possible choices are uncertain. For example, for 'frame type' (where the choice was '*in situ*', 'masonry', 'precast', 'steel' or 'timber'), there is a lack of consensus on the comparative costs, and it is impossible to ascertain a consistent ordering, in terms of cost, that would apply in all circumstances. Therefore, a binary input coding (yes/no) was applied to each possible choice, thus treating each such categorical variable as a series of binary variables.

Comparison of modelling techniques

A comparison of linear regression analysis and neural networks has been made elsewhere (Harding *et al.*, 1999b) and some preliminary analyses carried out (Harding *et al.*, 2000b). However, in this situation, the main advantage that neural networks offer is their ability to capture the nonlinearity that inevitably will exist between variables. Nonlinear regression techniques can also be used to account for nonlinearity, but have the disadvantage that the user must have detailed knowledge about the appropriate nonlinear relationship between the predictor variables and the mean values of the observations (Christensen, 1996). However, when applying neural networks, these relationships are determined implicitly by the model and therefore do not need to be specified.

Representation of cost (output variable)

Previously, cost models have often used the raw cost as the dependent variable. However, there are a number of assumptions implicit in this choice of variable. First, it is assumed that the standard deviation of error remains constant. That is to say, the cost of a small project can vary by the same monetary amount as a large project. This is highly unlikely to be the case. Further, regression model fitting minimizes the squares of the errors, so models developed using this technique will be inherently biased towards minimizing the errors for very large projects, where the errors are greatest. Therefore it is unlikely to be a good predictor of the cost of smaller projects. Given that the costs of projects in the data collected vary between £36 000 and £15 800 000, the influence of errors on the cost of the largest projects is several orders of magnitude more than those of the smallest projects, so the effect will be pronounced.

The second inherent assumption that is questioned is that the effects of any variable are best expressed as a fixed cost change. If, for example, the specification of the floor finishes changed to one of a higher cost, the cost of the building would be expected to rise. However, the cost of a small building would not be expected to rise by the same amount as the cost of a very large one, but as a proportion of the building size or cost.

These criticisms raise serious questions as to the meaningfulness of models produced by using raw cost as the predictor for a linear regression model. Therefore, three other possible models were tested.

Log of building cost

In order to address the problem of the large cost differences, a common solution is to model the log of the cost. This assumes that the log of cost is normally distributed, and that a change in any variable within the model will cause a proportionate change in cost. The distribution that corresponds to a normal distribution of the log of cost is the one whose mean is the project cost, and whose standard deviation is a fixed proportion of project cost. When the normal distribution is converted to raw cost for any project, it is a positively skewed distribution such that the peak of the probability density function is less than the mean. It might be argued that the skewed nature of this function could be a better representation of the possible variation in project cost than a true (unskewed) probability distribution, because generally there is more scope for the project costs to be much higher than expected rather than much lower.

Cost per m^2

The cost per m^2 is the cost predictor most used by quantity surveyors, as it provides a measure of cost that is essentially independent of building size. If this value were to be used in a regression model, then it would assume that any variation in project cost is proportional to the size of the building, rather than the cost. This may seem to be an unrealistic solution, because projects that are of a higher specification (and hence a higher cost per m^2) might be expected to show correspondingly higher variations in cost. However, it has the advantage of removing the understood linear relationship between GIFA and project cost from the model. This should allow the modelling to focus on other, less understood, influences on project cost.

Log of cost per m^2

The log of cost per m^2 makes the same assumptions as the log of cost, in that variations in project cost are proportional to the expected cost. However, it also provides a variable that is devoid of the linear relationship between cost and size, in the same way as the cost per m^2 output. Although this makes little difference to regression models, it could be useful in neural network modelling, if the correlation between cost and GIFA creates difficulties for those models that use the log of cost in learning the relationship between cost and other variables, because this could stop the neural network learning only this relationship to the detriment of others.

Factor analysis

Factor analysis was used to establish the underlying dimensions of the input variables, confirming that all items should be retained. Principal component extraction with varimax rotation was also used and, although this indicated that the true number of underlying dimensions lay between eight and nine factors, regression models created using the factor scores resulted in R^2 values less than those generated using the original input variables. Therefore a decision was reached to proceed using the original input variables.

Linear regression analysis

The first aim of the regression analysis was to develop a robust regression cost model that would make a useful benchmark against which neural network models could be measured. Second, it was desirable to identify those variables that demonstrated a strong linear relationship with the cost, to assist in the management of neural network training. The software used was SPSS 7.5 for Windows.

In order to create a predictive regression model, two methods were attempted: forwards and backwards modelling. Models to predict cost per m^2, log of cost per m^2 and log of cost were generated for each method. The number of statistically significant variables in each model varied between eight, in the forward stepwise log of cost model, and 14, in both the log of cost and log of cost per m^2 backward models. Throughout the models 19 different variables were used. A summary of the results is given in Table 2.

Five variables appeared in all six models: 'GIFA', 'function', 'duration', 'mechanical installations' and 'piling'. This suggests that these are the key linear cost drivers in the data. A further four variables appeared in five models: 'internal wall finishes', 'frame type', 'site access' and 'protective installations'.

The log of cost backwards models outperformed the other models by most of the percentage error measures. However, the differences between all the models were small. One of the reasons the log of cost and log of cost per m^2 models performed so well was that they incorporated more variables of whose inclusion in the model we could be confident, although they may not necessarily be the most useful variables in terms of which to model, because their values may be uncertain at the early estimating stage.

As the models exhibited similar performance, it might be more appropriate to consider the spread of error. Neural networks minimize error using the least-squares approach. As this can be sensitive to non-uniformity of standard deviation and non-normality of error, the spread and normality of error were assessed for each model by considering scatter plots of error against the value of the independent variable, all of which displayed the same tendency for the models to overestimate the cost of cheaper projects and underestimate the cost of more expensive projects. The fact that as much as 30% of the error appeared to arise from this suggests that some key drivers of building cost were not being represented adequately. This arises either from their non-inclusion or, more likely, from nonlinearities in the data, in which case it may be expected that a neural network will perform better.

Table 2 Linear regression model results

	Cost per m^2		Log of cost per m^2		Log of cost	
	R^2	MAPE	R^2	MAPE	R^2	MAPE
Forward	0.668	20.8%	0.648	20.0%	0.644	20.1%
Backward	0.666	21.7%	0.661	19.3%	0.661	19.3%

Neural networks

Three sets of models were developed, using:

- the five variables that were incorporated in all six of the linear regression models;
- the nine variables found in five of the six regression models; and
- all the variables for which data have been collected.

In addition, an optimum combination of variables was used, determined by using a combination of the forwards and backwards stepwise feature selection algorithm of a generalized regression neural network (GRNN) and a genetic algorithm (GA) global optimization search technique. Because theoretically a GA is better able to come up with an optimum combination of variables, and is a randomized process, this was repeated four times and the forwards and backwards algorithms were executed once each. By excluding all the variables that do not appear in the feature selection results, the model was reduced to a 25-variable model. However, as the performance of this model was poorer on both measures of performance than the all-variable model, suggesting that one or more significant variables had been omitted from the analysis, this approach was abandoned. The software used was Trajan Neural Network Simulator Release 4.0E.

In order to assess the best approach to the modelling, a number of different networks were tried: three- and four-layer multi-layer perceptrons (MLPs), radial basis functions (RBFs) and generalized regression neural networks (GRNNs). Of these alternatives, three-layer MLP networks (with one hidden layer) offered the best performance, in terms of the associated values of R^2 and mean absolute percentage error (MAPE).

To train the network, backpropagation, conjugate gradient descent, Levenberg Marquardt, quick propagation, delta-bar-delta and quasi-Newton training algorithms were evaluated. Of these, the ordinary back-propagation algorithm was the most efficient: it found solutions that were at least as good as the solutions found by other algorithms, and usually more quickly.

Once the structure and training method to be used had been established, three different function types were used as activation functions in the hidden and output layer: linear, logistic and hyperbolic. The hyperbolic function is approximately linear at zero, requiring the variables to be normalized and the weights set to very small values. This means that early in the training linear behaviour is assumed, while later the curved areas of the hyperbolic function are used to model the function. It was found that this technique yielded slightly better networks more quickly. The optimum number of nodes in the hidden layer was also investigated.

Five-variable model

The results of the best networks for the five-variable model gave R^2 values that were similar to but not as good as the regression analyses. However, the regression models were tested on the same data set used to create the model, while the neural network models were tested using a small (only 45 cases) independent test data set, which could cause the accuracy of the network to be misrepresented. The verification set is similarly prone to being unrepresentative. The process of training might, therefore, be terminated prematurely, when the model begins to lose its fit with the verification set, but before (or after) the model ceases to fit the real population. One solution to this problem is to use a 'voting system', involving the creation of a number of models, each of which uses different training, verification and test sets. The value of the output is then taken as the average of these models. Bias will exist within all the individual models, but provided all the projects in the data set are represented, other models will be biased in the opposite direction, producing an output that is closer to the mean of the data set, rather than just a small subset.

A voting system was trained for the cost per m^2 and log of cost per m^2 models. The values of R^2 were similar to those found when the voting system was not adopted, although the actual values for the individual models did vary significantly. This can be assessed by comparing the average R^2 values for each combination of training, verification and test sets using analysis of variance. The value of F obtained was 8.63, which is very highly significant, showing that the differences between the R^2 values have not arisen by chance, but from the fact that the test sets are not representative.

Examining the distribution of R^2 values, the best overall mean value is obtained by a network architecture containing two nodes in the hidden layer and, overall, the best values of both the mean and R^2 values found are for architectures with eight nodes in the hidden layer or less. However, these values have been obtained with only 20 models, and the differences are small.

Although the determination of the best architecture is difficult, because there is little significant difference between the performance of each, there are significant performance differences between different configurations of the training, verification and testing sets. This validates the approach of the voting system. However, although it is possible to average the results of the entire set, it is not possible to obtain accurate estimates of the true error using the test set, as each project has not been included in the test set enough times.

In order to permit more effective representation of each project, the number of networks used in the voting system was increased to 50, but there was no improvement over the best regression analysis. Nevertheless, the cost per m^2 neural network model did improve in respect of its linear counterpart. This implies that the model has encapsulated aspects of the relationships between the variables that the regression analysis has failed to do, suggesting that the neural network model is capable of improving upon the linear models where nonlinear relationships do exist.

Nine-variable model

Voting systems were trained in a similar way. These networks were created by 20 attempts at training for each of 15 different network architectures. The architectures were all three-layer MLP networks with between one and 15 nodes in the hidden layer. However, the results showed that with these models the neural networks awere failing to model even some of the linear relationships. It would appear that there are insufficient data to model this problem effectively. This is partly because only a small proportion of the relationships are nonlinear and most of the model can be explained using a linear model. Also, there are not sufficient data to permit good generalization of networks with an architecture adequate for modelling the nonlinearities. Generally, it is more difficult for neural networks to learn relationships that explain only a small part of the model's performance. This is because the variation is not large enough in proportion to the amount of unique error in the problem (error that cannot be explained by the variables).

All-variables model

This model was developed as a voting system trained to predict the cost per m^2. As the model involves a large number of variables, it was felt that the potential influence of bias would be larger than for the earlier models, and the size of the voting system was increased to 100 networks. The model demonstrated a significant improvement over the equivalent regression model. From the results of the earlier networks, it is known that a network of this size will find it very hard to model some of the more subtle linear relationships. Thus, the increase in performance observed implies that there are significant nonlinear relationships that the network is modelling.

The fact that earlier neural network models did not appear to find some of the nonlinear relationships has implications for the accuracy of this model. The number of variables has gone up, making it harder for the neural network to model some of these weaker relationships. Therefore, the model is almost certainly not modelling some of these relationships, which has a detrimental effect on its accuracy. In addition to this, it is also likely that the modelling of the nonlinear relationships also lacks some accuracy for the same reasons. Therefore, although it already outperforms the regression model, clearly more data would produce an improvement in the model's performance.

Results

The results obtained applying the methods outlined above are given in Table 3, which also shows (in parentheses), where appropriate, the results obtained for regression models developed using the same variables, so that a direct comparison can be made between models obtained by the two techniques.

The best model obtained was a neural network model using all 41 variables and a voting system using 100 networks; this gave an R^2 value of 0.789 and a MAPE of 16.6%, which is an improvement upon results obtained from regression analyses generally and the 'best' regression model specifically (R^2 of 0.661 and MAPE of 19.3% for the backward log cost model). Where linear regression and neural network models have been developed using the same variables, neural networks always outperform their regression counterparts, and the best linear regression models always outperform those developed for direct comparison with neural networks.

These results compare favourably with past research that has shown that traditional methods of cost estimation are less accurate, as evidenced by reported values of MAPE between 20.8% (Skitmore *et al.*, 1990) and 27.9% (Lowe, 1996).

Conclusions

The two approaches to modelling, predicting the cost per m² and the log of cost per m², achieve a similar performance but with subtle differences. Predicting the cost per m² tends to produce a model with a higher R^2 value in cost per m² terms. However, the log model yields lower values of MAPE. This is a function of the fact that the log model explicitly minimizes proportional differences, whereas the untransformed cost per m² model minimizes the square of the error on the cost per m². Thus the model selected should be based upon whether the user wants accuracy in proportional or cost per m² terms.

The overall results have significant implications for the assumptions on which the research is based. It was assumed that there were definite benefits to using a neural network approach, as this should be capable of modelling the nonlinear relationships in the data. While the models presented above may not be much more accurate than current cost estimation practice, they do show that there are nonlinearities in the data and that neural networks are capable of modelling them. It is believed that further analyses will lead to better models, and that there is evidence that the inclusion of more data would yield significant improvements

Table 3 Neural network model results

	Cost per m²		Log of cost per m²	
	R^2	MAPE	R^2	MAPE
Five-variable model (without voting system)	0.564	20.3%	0.586	18.3%
Five-variable model (with voting system)	0.636 (0.556)	21.9% (23.6%)	0.641 (0.618)	19.6% (22.2%)
Nine-variable model (with voting system)	0.688 (0.617)	20.1% (21.5%)	0.685 (0.649)	18.6% (20.7%)
All-variables model (with voting system)	0.789 (0.683)	16.6% (27.7%)		
Reduced variables model (with voting system)	0.665	22.7%		

in the accuracy that could be achieved by neural network modelling. These analyses are currently being undertaken.

Acknowledgements

The authors acknowledge, with great gratitude: the support of the Engineering and Physical Science Research Council, who funded the project; our professional collaborators, E C Harris, Symonds and Tweeds; the BCIS for their contribution during the full scale study; and the many other quantity surveying and project management practices who provided data.

References

Christensen, R. (1996) *Analysis of Variance, Design and Regression: Applied Statistical Methods*. Chapman & Hall, London.

Corbett, P. and Rowley, R. (1999) The use of BCIS elemental cost data by quantity surveyors as part of cost planning techniques: the practitioners' perspective, in *Proceedings of the 15th Annual ARCOM Conference*, pp. 465–72.

Duff, A. R., Emsley, M. W., Gregory, M., Lowe, D. J. and Masterman, J. (1998) Development of a model of total building procurement costs for construction clients, in *Proceedings of the 14th Annual ARCOM Conference*, pp. 78–87.

Elhag, T. and Boussabaine, A. (1998) An artificial neural system for cost estimation of construction projects, in *Proceedings of the 14th Annual ARCOM Conference*, pp. 219–26.

Elhag, T. and Boussabaine, A. (1999a) Tender price estimation: neural networks vs. regression analysis, in *Proceedings of the RICS Construction and Building Research Conference*, University of Salford, pp. 114–23.

Elhag, T. and Boussabaine, A. (1999b) Evaluation of construction cost and time attributes, in *Proceedings of the 15th Annual ARCOM Conference*, pp. 473–80.

Ferry, D. J., Brandon, P. S. and Ferry, J. D. (1999) *Cost Planning of Buildings*, 7th Edn, Blackwell, Oxford.

Harding, A., Lowe, D. J., Emsley, M. W., Hickson, A. and Duff, A. R. (1999a) The role of neural networks in early stage cost estimation in the 21st century, in *Proceedings of the RICS Construction and Building Research Conference*, University of Salford, pp. 161–8.

Harding, A., Lowe, D. J., Emsley, M. W., Hickson, A. and Duff, A. R. (1999b) Implementation of a neural network for the comparison of the cost of different procurement approaches, in *Proceedings of the 15th Annual ARCOM Conference*, pp. 763–72.

Harding, A., Lowe, D. J., Emsley, M. W., Hickson, A. and Duff, A. R. (2000a) Implementation of a neural network model for the comparison of the cost of different procurement approaches, in *Information and Communication in Construction Procurement*, CIB W92, pp. 269–80.

Harding, A., Lowe, D. J., Emsley, M. W., Hickson, A. and Duff, A. R. (2000b) The cost of procurement: a neural network approach, in *Proceedings of CIT2000: Construction Information Technology*, CIB W78, pp. 428–38.

Lowe, D. J. (1996) Experiential learning in design cost estimating. Ph.D. thesis, UMIST.

Newton, S. (1991) An agenda for cost modelling research. *Construction Management and Economics*, **9**(2), 97–112.

Skitmore, M., Stradling, S., Tuohy, A. and Mkwezalamba, H. (1990) *The Accuracy of Construction Price Forecasts*, University of Salford, Manchester.

Tarassenko, L. (1998) *A Guide to Neural Computing Applications*, Arnold, London.