A CONSTRUCTION COST FLOW RISK ASSESSMENT MODEL

Henry A. Odeyinka[1], John G. Lowe[1] and Ammar Kaka[2]

[1] *School of the Built & Natural Environment, Glasgow Caledonian University, Glasgow G4 0BA*
[2] *Department of Building Engineering & Surveying, Heriot-Watt University, Edinburgh EH14 4AS*

Various 'short cut' approaches have been adopted in cash flow forecasting ranging from the statistical, mathematical, simulation and the use of artificial intelligence techniques. However, the majority of these approaches failed to consider the issues of risks and uncertainties inherent in construction. As such, a wide variation is observable between the predicted cash flow profile and the actual. This study attempts to model the variation between predicted and actual cost flow due to inherent risks in construction. Data were obtained through questionnaire survey and empirical data collection. Contractors on individual projects were requested to score on a Likert type scale, the extent of occurrence of each identified risk variable that resulted in the variation between the predicted and actual cost flow profiles. An analysis of the responses, using ranking of the mean response enabled the study to focus on the most significant risk variables. The impact of these risk variables on cost flow forecast was then investigated by collecting data on predicted and actual cost flow from completed construction projects in order to determine their variation. Neural network was employed using the back propagation algorithm to develop the cost flow risk assessment model. The developed model was tested on 20 new projects with satisfactory predictions of variations between the forecast and actual cost flow at 30, 50, 70 and 100% completion stages.

Keywords: cost flow, contractor, modelling, neural network, risk and uncertainty.

INTRODUCTION

Cash flow forecasts are of great importance to construction contractors as well as the client to prevent the unsavoury consequences of liquidation and bankruptcy. However, an accurate forecast of construction cash flow has been a difficult issue due to risks and uncertainties inherent in construction projects. According to Flanagan and Norman (1993: 22), the environment within which decision making takes place can be divided into three parts: certainty, risk and uncertainty. According to them, certainty exists only when one can specify exactly what will happen during the period of time covered by the decision. This, they concluded of course does not happen very often in the construction industry. Bennett and Ormerod (1984) also concluded that an important source of bad decisions is illusions of certainty. They submitted that uncertainty is endemic in construction and needs to be explicitly recognised by construction managers. However, in the last three decades, many deterministic cash flow forecasting models have been developed (Odeyinka 2001) which did not take into consideration, the issue of risks inherent in construction. As such, considerable variations were observable between the modelled and actual cash flow. In an attempt to increase forecasting accuracy, nomothetic and idiographic epistemology were

[1] E-mail: H.Odeyinka@gcal.ac.uk

Odeyinka, H A, Lowe, J G and Kaka, A P (2002) A construction cost flow risk assessment model. *In:* Greenwood, D (Ed.), *18th Annual ARCOM Conference*, 2-4 September 2002, University of Northumbria. Association of Researchers in Construction Management, Vol. 1, 3-12.

employed in previous researches. Also, deterministic and stochastic methodologies were used as well as the schedule-based or cost profile method (Kenley 2001). Value and cost flow approaches were also employed (Kaka 1999). Majority of the cash flow forecasting models employed the value approach. However, Kaka and Evans (1998) demonstrated that with very detailed classification of construction projects, it was very difficult to model the value curve accurately. Kaka and Price (1993) however were able to use the cost approach to classify and model construction cost flow more accurately, relative to previous models. Given this fact and also given the fact that variation between the modelled and actual cost flow is inevitable due to inherent risks in construction, this study, which is an on-going research therefore employs the cost approach to assess the impacts of risks on construction cost flow profile.

RISKS IN CASH FLOW FORECASTING

The major problem that construction managers encounter in making financial decisions involves both the uncertainty and ambiguity surrounding expected cash flows (Eldin, 1989). In the case of complex projects, the problem of uncertainty and ambiguity assumed even greater proportion because of the difficulty in predicting the impact of unexpected changes on construction progress and consequently, on cash flows. The uncertainty and ambiguity are caused not only by project-related problems but also by the economical and technological factors (Laufer and Coheca, 1990). Lowe (1987) maintained that the factors responsible for variation in project cash flow could be grouped under five main headings of contractual, programming, pricing, valuation and economic factors. Kaka and Price (1993) and Kaka (1996) in developing a model for cash flow forecasting identified other risk factors affecting cash flow profiles to include estimating error, tendering strategies, cost variances and duration overrun. Khosrowshahi (2000) also identified other risk factors that impact on cash flow to include delay payment and difficulty in obtaining the right amount of funds at reasonable interest rates.

Kenley and Wilson (1986) maintained that individual variation between projects' cash flow profile is caused by a multiplicity of factors, the great majority of which can neither be isolated in sample data, nor predicted in future projects. According to them, some existing cash flow models hold that generally two factors, date and project type, are sufficient to derive an ideal construction project cash flow curve. Such convenient division according to them ignores the complex interaction between such influences as economic and political climate, managerial structure and actions, union relations and personality conflicts. Many of these factors have been perceived to be important in related studies such as cost, time and quality performance of building projects (Ireland, 1983). According to Kenley and Wilson (1986), models, which ignore all these factors in cash flow research, must be questioned. Fig. 1 shows a typical variation between the forecast and actual cash flow occasioned by risk impacts. While this variation has been identified in literature, the impact of risks inherent in construction resulting in the variation has not been investigated. This then is the concern of this study.

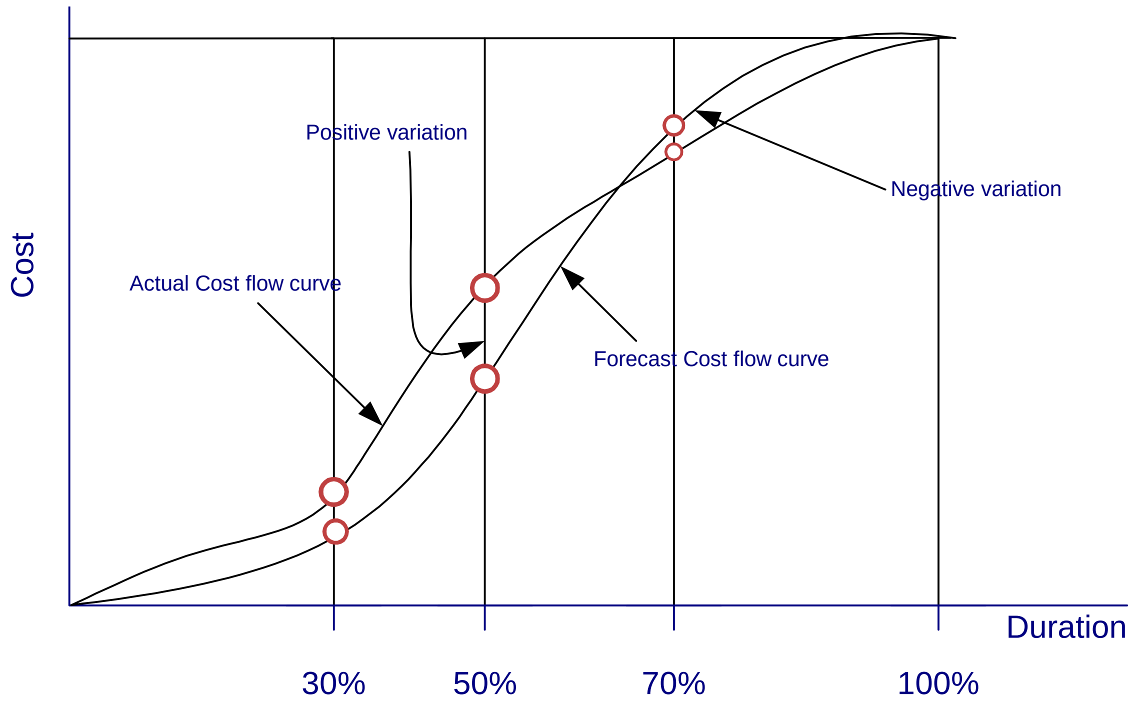

Fig 1:A typical variation between forecast and actual cost flow curves at 30%, 50%, 70% and 100% completion

DATA AND METHODOLOGY

Data were obtained through a combination of questionnaire survey and collection of empirical data from archival sources. Two sets of questionnaires were employed in data collection. The first set identified 25 risk factors from literature and from discussion with other researchers in construction cash flow as well as from discussion with construction practitioners. These factors were perceived to have potential impact on cost flow forecast. The questionnaire was then administered on a project by project basis to 350 randomly selected small, medium and large-scale contractors. A reminder letter subsequently followed this. In all, 96 responses fit for analysis were received, which represents a 27.4% response rate. The contractors were asked to score on a Likert type scale of 0-5, the extent of occurrence and perceived impacts of the identified risk factors on a recently completed building project. This approach has been detailed in a related pilot study (Odeyinka and Lowe 2000, 2001a). Mean response analysis was performed on the responses. This was ranked in order to determine the significant risk factors to focus on. Table 1 summarises the result of this analysis. From the table, the first 11 high ranking risk factors, which also happen to have high-ranking impacts were selected so as to assess their impacts on variation between the forecast and actual cash flow profile. The cut off point was determined using the critical point of 3.0 on the 0-5 Likert scale. This was applied to the risk impact mean score. The reason being that some risk factors can be low in occurrence but high in impact. Using the 11 high ranking risk factors, another questionnaire was administered on a project by project basis to medium and large-scale construction contractors. The questionnaire asked them to score on a scale of 0-5 the extent of occurrence of the identified risk factors on a recently completed or an on-going building project. It also requested them to estimate the percentage variation between their forecast cost flow and actual cost flow on the project at 30, 50, 70 and 100%

completion stages. The rationale for choosing these time intervals is that previous research (Kaka, 1999) has shown that majority of variability between cost flow curves occurred between the 30% and 70% portions of the curves. The 100% point was also selected in order to track down possible cost overruns at the end of the contract period.

Out of the 400 questionnaires administered, only 50 were returned and fit for analysis. This represents a 12.5% response rate. Empirical data were also obtained from the archives of a construction company regarding the forecast and actual cost flow of 10 recently completed building projects. The project quantity surveyor on each project was also requested to complete a questionnaire regarding the extent of occurrence of the identified risks on the project. Variations between the forecast and actual cost flow at 30, 50, 70 and 100% completion stages were later calculated to get a complete data set. Out of the 50 data set obtained from the second set of questionnaire survey, 40 were used to develop a cost flow risk assessment model while the remaining 10 and the other 10 data set obtained from a construction company were used for testing and validating the model. The model was developed using the artificial neural network (ANN).

An artificial neural network is an information processing system that has certain performance characteristic in common with biological neural networks. Artificial neural networks have been developed as generalisations of mathematical models of human cognition or neural biology. A neural network is characterised by its pattern of connections between the neurons (called its *architecture*), its methods of determining the weights on the connections (called its *training*, or *learning, algorithm*), and its *activation function* (Fausett 1994).

It has been proved that neural network can solve problems with multi-attributes better than conventional methods (Masters 1993). Neural networks are suited to such problems because of their adaptivity owing to their structure; i.e. non-linear activation functions. Unlike expert systems and traditional modelling methods where knowledge is made explicit in the form of rules, neural networks generate their own rules by learning from examples (Gallant 1993). The problem of learning in neural network is simply that of finding a set of connection strengths that allow the network to carry out the desired computation (Boussabaine 1998). The learning method used in this study is the back propagation (BP) which is the most widely and successfully used algorithm in neural networks.

The main mechanism in a back propagation network according to Boussabaine (1998) is to propagate the input forward through the layers to the output layer and then to propagate the errors back through the network from the output layer to the input layer. The input data in this study were the identified significant risk factors while the output data were the estimated variation at 30, 50, 70 and 100% completion stages. A BP neuron transfers its inputs as follows:

$$\text{Output (node)}_{\text{i}} = \sigma\, [\sum w_{ij} x_j(\text{t}) - \beta_{\text{I}}]$$

Where σ is the sigmoid function, w_{ij} is the strength of the connection (weight) from node j to node i, x_j is the output value of node j; and β_{I} is the node threshold value. As such, when a neuron is activated, the new output is equal to the sigmoid function of the sum of

Table 1: Perception of risk occurrence and impact on cost flow forecast

Risk Factors	Risk occurrence mean score	Rank	Risk impact mean score	Rank
Changes to initial design	3.32	1	3.72	1
Inclement weather	3.00	2	3.72	1
Variation to works (AI)	2.95	3	3.55	3
Labour shortage	2.81	4	3.53	4
Production target slippage	2.70	5	3.47	5
Delay in agreeing variation/dayworks	2.62	6	3.28	6
Delay in settling claims	2.59	7	3.20	7
Problems with foundations	2.46	8	3.15	8
Underestimating project complexity	2.41	9	3.10	9
Estimating error	2.24	10	3.01	10
Under valuation	2.24	10	3.00	11
Delay in payment from client	2.08	12	2.35	12
Shortage of key materials	2.08	12	2.30	13
Delays in interim certificates	2.03	14	2.27	14
Delay in retention release	1.97	15	2.16	15
Inflation	1.86	16	2.03	16
Compliance with new regulations	1.78	17	1.95	17
Subcontractor's insolvency	1.70	18	1.81	18
Changes in interest rates	1.68	19	1.76	19
Shortage of key plant items	1.68	19	1.62	20
Access to funds at reasonable interest rate	1.46	21	1.62	21
Archaeological remains	1.46	21	1.54	22
Changes in currency exchange rates	1.35	23	1.41	23
Civil disturbances	1.24	24	1.35	24
Labour strikes	1.19	25	1.32	25

the products of the weights and the activities of the input wires (connections), minus the threshold of the node. The sigmoid function is defined by the following equation:

$$\sigma(x) = \frac{1}{1 + \exp(-x)}$$

The global error function used to propagate the error back through the network is

$$E = \sum_i (d_i - o_i)^2$$

Where d_i is the desired output and o_i is the actual output produced by the network. The main objective is to maximise this function, i.e. to change the weights of the system in proportion to the derivative of the error with respect to the weights. The error correction learning procedure is simple in conception and is as follows: during training, input is provided for the network and flows through the network generating a set of values on the output neurons. Then, the actual output is compared with the desired target, and a match is computed. If the output and the target match, no change is made to the net. However if the output differs from the target a change must be made to some of the connections. This procedure works remarkably well on a variety of problems (Masters 1993).

DEVELOPMENT OF THE ANN MODEL

Neural network based modelling process according to Ogunlana *et. al.* (2001) involves five main aspects. These are (1) data acquisition, analysis and problem representation; (2) architecture determination; (3) learning process determination; (4) training of the network; and (5) testing of the trained network for generalisation evaluation.

As previously stated 40 data sets were used in the model development. The scores attached to the 11 identified significant risk factors were used as the input for the neural network. The estimates of percentage variation between the forecast and actual cost flow at 30, 50, 70 and 100% completion stages were used as the output. These two pairs of data set constitute the training set for the neural network model. The number of input data and output data also suggested that 11 input neurons and 4 output neurons would be good network architecture to start with. After a number of trials and errors, the network was found to stabilise with 12 hidden nodes. Thus, the model employed 11- 12- 4 back propagation architecture as shown in Fig 2.

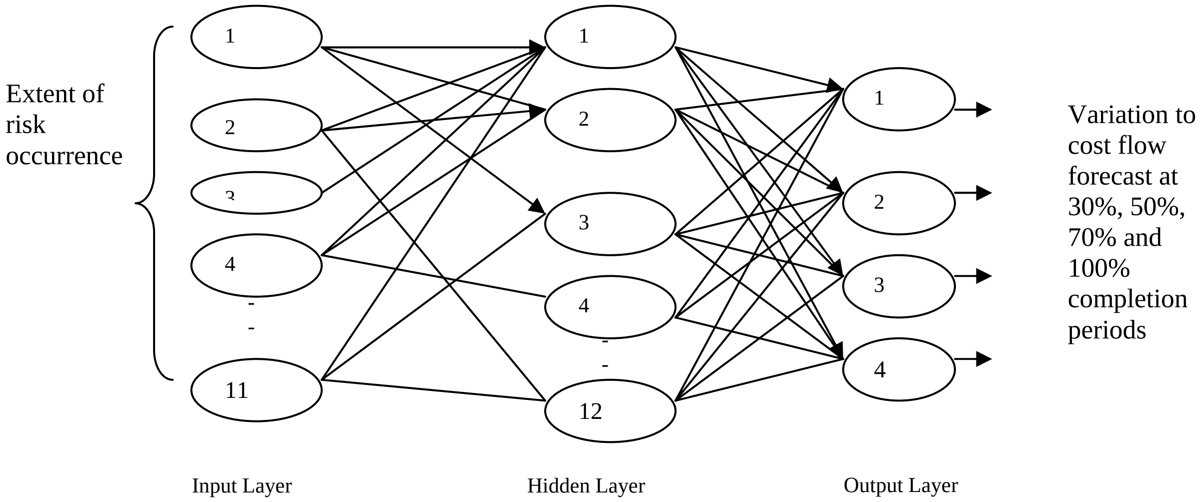

Fig. 2: Artificial network architecture employed for assessing risk impacts on cost flow forecast cost flow forecast

Having determined the network architecture, the input output pairs for the 40 data sets used in training the network for model development were fed into the neural network software employed. Using the sigma transfer function as defined in the previous section and back propagation algorithm, the neural network tried to establish in a black box manner, the relationships between the inputs and outputs during training. A screen dump of the training pairs and the networks generated is shown in the Appendix. The learning rate η for the best network was found after trial and error to be 0.7 with 32001 training cycles. Training error was set to be reduced to 0.001. The outputs produced by the network and the desired outputs of the training samples were then compared using the mean squared error (RMS). This according to Fausett (1994: 3-7) is computed using the formula:

$$RMS = \sqrt{\{\frac{1}{n}\sum_{i=1}^{n}(t_i - o_i)^2\}}$$

Where t_i and o_i are actual and predicted variation at completion period *i* and *n* is the number of periods. After 32001 training cycles, the RMS for the training samples was found to be 0.007, which is not too far from the target of 0.001. This suggests that the system has learned the relationship between the inputs and outputs and can also generalise from data. The network architecture together with the associated weight matrix is shown in the Appendix. This now becomes a model for assessing risk impacts on construction cost flow forecast.

In order to test the forecasting adequacy of the model, the remaining 20 data set were used. Input data regarding the extent of risk occurrence were entered and the network requested to forecast the percentage variation at 30, 50, 70 and 100% completion periods. The outputs are shown in the screen dump in the Appendix. The validation tests performed on the networks were a comparison between the predicted and the actual percentage variation obtained from the test data at the desired completion periods. The statistical verification method employed was the relative mean absolute deviation (Rel. MAD). This was chosen as it is a unitless quantity and allows deviations to be measured in absolute terms. The formula for this measure is given as:

$$\mathrm{Re}\,l.MAD = \frac{1}{n}\sum_{i=1}^{n}\frac{|\,t_i - o_i\,|}{t_i}$$

Where t_i and o_i are actual and predicted variation at completion period i and n is the number of periods. Table 2 shows the actual and predicted variations at 30, 50, 70 and 100% completion stages. A cursory look at the prediction error shows that the deviations are not too much. However, for a better perception of the error, the Rel. MAD is calculated as stated earlier. The Rel. MAD measures for each of the 20 data sets used in testing the model are shown in Table 2 and Fig 3 also depicts the Rel. MAD measures for the 20 test projects.

Table 2: Network prediction performance using relative mean absolute deviation measurement

Proj. No.	30% Completion			50% Completion			70% Completion			100% Completion			Rel. MAD
	Actual	Predicted	Error	Actual	Predicted	Error	Actual	Predicted	Error	Actual	Predicted	Error	
1	0.12	0.16	-0.04	0.08	0.07	0.01	0.08	0.03	0.05	0.05	0.01	0.04	0.471
2	0.15	0.20	-0.05	0.12	0.15	-0.03	0.10	0.05	0.05	0.05	0.09	-0.04	0.471
3	0.20	0.26	-0.06	0.26	0.40	-0.14	0.19	0.44	-0.25	0.23	0.30	-0.07	0.615
4	0.21	0.27	-0.07	0.19	0.23	-0.04	0.25	0.31	-0.06	0.04	0.07	-0.03	0.372
5	0.18	0.22	-0.04	0.19	0.18	0.01	0.12	0.16	-0.04	0.05	0.01	0.04	0.352
6	0.19	0.12	0.07	0.43	0.37	0.06	0.17	0.16	0.01	0.07	0.02	0.05	0.320
7	0.18	0.15	0.03	0.26	0.25	0.01	0.07	0.10	-0.03	0.01	0.05	-0.04	1.158
8	0.36	0.30	0.06	0.24	0.23	0.01	0.17	0.18	-0.01	0.01	0.01	0.00	0.067
9	0.25	0.29	-0.04	0.30	0.34	-0.04	0.30	0.36	-0.06	0.05	0.01	0.04	0.323
10	0.05	0.07	-0.02	0.10	0.10	0.00	0.15	0.07	0.08	0.10	0.03	0.07	0.408
11	0.25	0.28	-0.03	0.25	0.32	-0.07	0.20	0.06	0.14	0.10	0.01	0.09	0.500
12	0.10	0.06	0.04	0.15	0.04	0.11	0.20	0.19	0.01	0.20	0.29	-0.09	0.408
13	0.15	0.25	0.10	0.25	0.29	-0.04	0.35	0.30	0.05	0.20	0.28	-0.08	0.342
14	0.20	0.28	-0.08	0.15	0.18	-0.03	0.10	0.06	0.04	0.05	0.01	0.04	0.450
15	0.15	0.20	-0.05	0.10	0.14	-0.04	0.10	0.08	0.02	0.05	0.01	0.04	0.433
16	0.15	0.16	-0.01	0.20	0.25	-0.05	0.20	0.16	0.04	0.10	0.02	0.08	0.329
17	0.20	0.15	0.05	0.30	0.37	-0.07	0.45	0.43	0.02	0.10	0.12	-0.02	0.182
18	0.15	0.14	0.01	0.20	0.26	-0.06	0.20	0.10	0.10	0.15	0.18	-0.03	0.267
19	0.20	0.30	-0.10	0.25	0.34	-0.09	0.35	0.42	-0.07	0.05	0.01	0.04	0.465
20	0.05	0.03	0.02	0.10	0.04	0.06	0.15	0.15	0.00	0.20	0.16	0.04	0.300

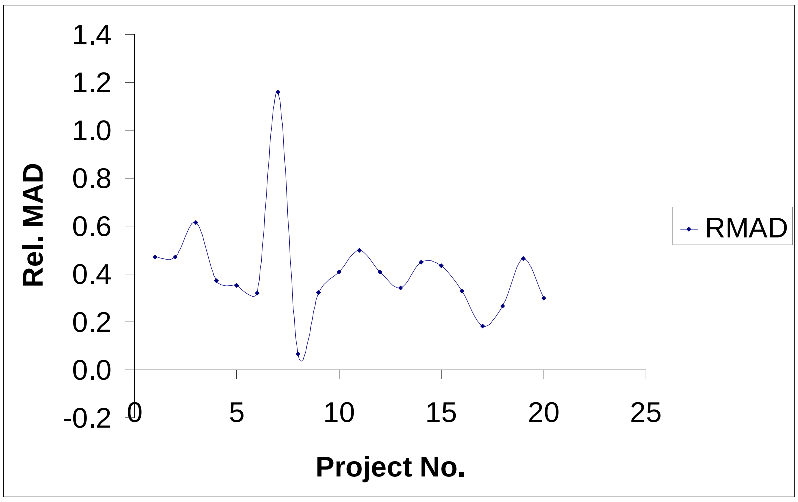

Fig. 3: Relative mean absolute deviation (Rel. MAD) for test projects

It is obvious from Table 2 and Fig. 3 that the Rel. MAD clusters between 0.18 and 0.5, which is considered good enough for the type of data, involved. However, project 7 has a very high Rel. MAD of 1.16, followed by project 3 with Rel. MAD of 0.62. The data sets for these 2 projects are from the empirical data collected from a construction company. While this may be considered as an outlier compared to other test samples, it is also possible that there may be an error of judgement in scoring the risk variables as carefully scored variables are very important for accurate predictions. Further work would also be carried out to determine the accuracy of the model using the SDY as proposed by Kenley and Wilson (1986), Kaka (1999), etc. The SDY measures the standard deviation about the estimate of Y. Results would also be compared with accuracy achieved in previous models.

CONCLUSION

This paper identified the issues of risks in construction as a major cause of the observed variations between the forecast and actual cost flow. It then went ahead to identify some significant risk factors which potentially impact on cost flow forecast. These risk factors were focused upon to examine their impacts on the variation between the forecast and actual cost flow. The data obtained from 40 building projects were used to develop an artificial neural network model to assess risk impacts on cost flow forecast. Results from testing and validating data from 20 other building projects showed that an artificial neural network model could be useful in mapping and modelling the complex interaction between risks in construction and the usual variation observed between the forecast and actual cost flow.

REFERENCES

Bennett, J. and Ormerod, R.N. (1984) Simulation applied to construction projects, *Construction Management and Economics*, **2** (3), 225-63.

Boussabaine, A.H. and Kaka A.P. (1998) A neural networks approach for cost flow forecasting. *Construction Management and Economics*, **16**, 471-479.

Eldin, N. (1989) Cost control systems for PMT use. *Transactions of the AACE*, F3.1-F3.5.

Evans, R.C. and Kaka, A.P. (1998) Analysis of the accuracy of standard/average value curves using food retail building projects as case studies. *Engineering, Construction and Architectural Management*, **5** (1) 58-67.

Fausett, L. (1994) Fundamentals of Neural Networks: Architectures, Algorithms and Applications, Prentice Hall, Englewood Cliffs, NJ.

Flanagan, R. and Norman, G. (1993) *Risk Management and Construction*. Blackwell Science, London.

Gallant, S. I. (1993) *Neural Network Learning and Expert Systems,* MIT Press, Cambridge, MA.

Ireland, V. (1983) *The role of managerial actions in the cost, time and quality performance of high rise commercial building projects*. Unpublished PhD Thesis, University of Sydney, Australia.

Kaka, A P (1999): The development of a benchmark model that uses historical data for monitoring the progress of current construction projects. *Engineering, Construction and Architectural Management* **6**(3), 256-266.

Kaka, A.P (1996) Towards more flexible and accurate cash flow forecasting. *Construction Management and Economics*, **14**(1), 35-44.

Kaka, A.P. and Price, A.D.F. (1993) Modelling standard cost commitment curves for contractors' cash flow forecasting. *Construction Management and Economics*, **11**, 271-283.

Kenley, R. and Wilson, O. (1986) A construction project cash flow model – an ideographic approach. *Construction Management and Economics,* **4**, 213-232.

Kenley, R (2001) In-project end-date forecasting: an idiographic, deterministic approach,using cash-flow modelling. *Journal of Financial Management of Property and Construction,* **6** (3), 209-216.

Khosrowshahi, F. (2000) A radical approach to risk in project financial management. *In:* Akintoye, A (ed.) *Procs 16th Annual ARCOM Conference,* Glasgow Caledonian University, September 6-8, 547-556.

Laufer, A. and Coheca, D. (1990) Factors affecting construction planning outcomes. *Journal of Construction Engineering and Management*, **116**(6), 135-56.

Lowe, J.G. (1987) Cash flow and the construction client – a theoretical approach, in Lansley, P.R. and Harlow, P.A. (Eds.) *Managing Construction Worldwide*, Spon, London, **1**, 327-336.

Masters, T. (1993) *Practical Neural Networks Recipes in C++*, Academic Press, London.

Odeyinka H.A. & Lowe J.G. (2001) An analysis of the impacts of risks and uncertainties on construction cash flow forecast. *In:* Kelly A. & Hunter K. (Eds.) *Procs RICS Foundation Construction and Building Research Conference* (COBRA), Glasgow Caledonian University, September 3-5, 2001. 37-47.

Odeyinka H.A. & Lowe J.G. (2000) An assessment of risk factors involved in modelling cash flow forecast. *In:* Akintoye, A (ed.) *Procs 16th Annual ARCOM Conference,* Glasgow Caledonian University, September 6-8, 557-565.

Ogunlana S.O, Bhokha, S and Pinnemitr, N. (2001) Application of artificial neural network (ANN) to forecast construction cost of buildings at the pre-design stage. *Journal of Financial Management of Property and Construction*, **6** (3), 179-192.

APPENDIX: A SCREEN DUMP OF ANN MODEL ARCHITECTURE, WEIGHT MATRIX, TRAINING AND TEST DATA