A COMPARISON OF MULTIPLE REGRESSION AND ARTIFICIAL NEURAL NETWORKS APPROACHES TO IDENTIFYING SIGNIFICANT CONTRIBUTORS TO PROJECT PERFORMANCE LEVELS

by

Sunil M. Dissanayaka and Mohan M. Kumaraswamy

Abstract

This study attempts to identify and model the relationships between procurement and non procurement related (intervening) variables on the one hand; and project performance measures on the other. Multiple Regression (MR) analysis suggested that non procurement related variables are more associated with the time and cost over-runs in Hong Kong building projects than procurement related variables. The models for predicting time and cost over-runs, developed by MR, were compared with models based on an Artificial Neural Network (ANN) developed using the same data. The results suggest that the reliability of estimates by the ANN - based models is much higher than that of the MR-based models for the training data sets. Predictions for a testing set however, were found to be more reliable with MR-based models.

1.0 Background and Introduction

Identifying the principal factors contributing to project performance would provide valuable guidelines to optimise the achievement of client objectives through proper planning, controlling, co-ordination and timely decision making. Naoum [1994] suggested that the relevant experience of the building team significantly influenced the time and cost over-runs, as well as the quality achieved in a project. On the other hand, pre-construction time and unit cost were found to be more dependent on whether the client had selected a management or a traditional form of a contract. Masterman [1992] asserted that one of the principal reasons for the construction industry's poor performance is the inappropriateness of the procurement systems that have been chosen. Rowlinson [1989], and Walker [1994, 1995] concluded that management, organisational and contractual variables are highly associated with overall project performance. Ireland [1985] identified the effects of specific managerial actions on the achievement of the objectives of reducing cost, and time and increasing quality.

On the basis of the foregoing, the experiences of the authors and preliminary studies, the potential factors influencing project performance were grouped into (a) procurement related factors: work packaging, payment method, selection method, contract conditions and functional grouping; and (b) non procurement related factors: project characteristics, project team characteristics and their achieved performance levels, as well as external conditions. However, the procurement and non procurement related factors are not entirely mutually exclusive. Procurement arrangements lead to different patterns of responsibilities and relationships between clients and the various other parties involved in the building process [Nahapiet and Nahapiet, 1985].

The scope of a procurement system can be defined based on the CIB W92 definition as "the framework within which construction is brought about, acquired or obtained" [Sharif and Morledge, 1996]. As a total system, a procurement system can be considered to be an amalgamation of sub-systems of work packaging, functional grouping, payment modality, selection methodology and contract conditions. Using the hierarchy of possible procurement sub-systems, and options developed by Kumaraswamy and Dissanayaka [1996] and expanded further as in Figure 1, different procurement systems can be assembled by selecting and integrating options from each sub system. The most appropriate procurement system for a project should be selected to match the project scenario, because there is no "best procurement route". The type of client and his priorities for the particular project generally lead to some route being better than others [Turner, 1990].

It was therefore decided to incorporate all procurement sub-systems and intervening variables into one comprehensive model. It was envisaged that identifying the relative strengths of the linkages between pro-

Eng.(Dr) Mohan M Kumaraswamy - BSc Eng, MSc, PhD, CEng, FICE, FIE(SL), FCIOB, FHKIE, MASCE, MIE(AUS), ACIArb Associate Professor at the University of Hong Kong

Eng. Sunil M. Dissanayaka - BSc Eng, AMIE(SL), A Research Student at the University of Hong Kong

curement related variables and project performance would help to provide valuable guidelines during the procurement system selection process. A research project was thus formulated to explore those linkages.

This paper: (a) outlines the methodology and principal methods used during this research and (b) briefly summarises the data analysis, results and conclusions.

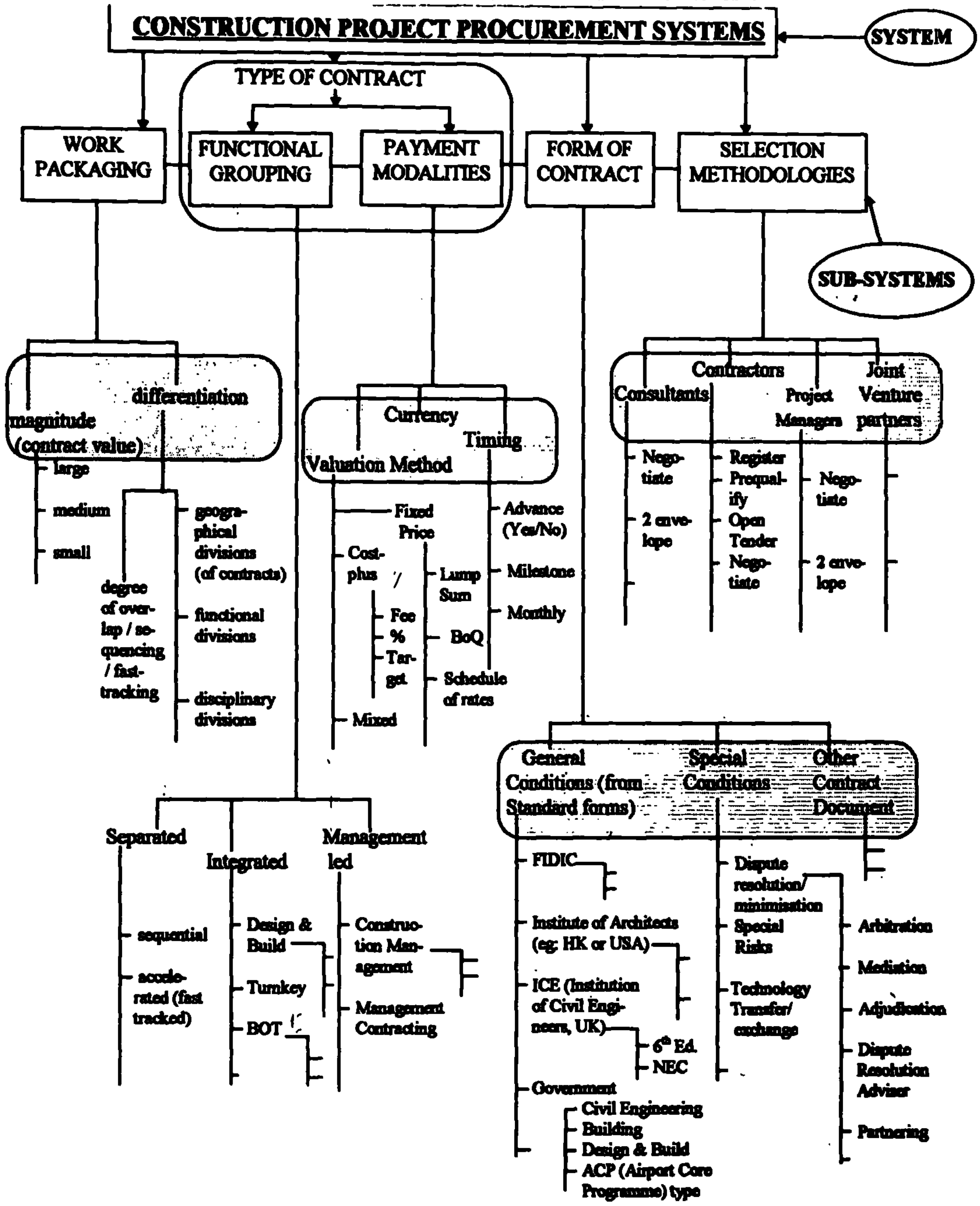

Figure 1: A Hierarchy of Procurement Options

NOTES:

▭ contains the 'sub-sub-systems' within each sub-system (ie indicates a break-down of the sub-system)

Each sub system has series of options as indicated above
All possible 'sub-sub systems' and options are not indicated here

Figure 1: A Hierarchy of Procurment Options

2.0 Research objective and strategy

The main objective of the research is to identify the strengths of association of procurement and non procurement related variables with project performance (of time and cost) in Hong Kong building projects. The following research strategy was derived from this main objective:

To identify and select particular variables that may be significantly related to project performance

To develop time and cost over-run models using such variables (procurement sub-systems and non-procurement related or 'intervening' variables) influencing project performance (in terms of meeting cost and time targets), using an appropriate method (eg: Multiple Regression [MR] analysis of quantitative and qualitative data obtained).

To use another method (eg: Artificial Neural Networks [ANN]) to develop time and cost over-run models using the significant factors identified from the previous stage.

To carry out a comparative study to compare the prediction capabilities of the two models.

3.0 Methodology

3.1 Identifying variables

The conclusions of previous researchers Rowlinson [1989], Naoum and Musthapha [1994], Walker [1994, 1995], Hashim [1996], Ashley et al. [1987], Hughes [1989] and Liu [1995], as well as initial interviews with industry practitioners were useful in identifying variables which may have some influence on project performance. These were grouped under the macro variables, each of which then 'covered' a large number of micro variables. All the selected micro variables may not be of the same importance in every project. The importance may vary with the client's objectives, priorities, project conditions, constraints and complexities and the quality of the project team. The selected variables were grouped into procurement and non-procurement related variables. Appendices 1 and 2 show these variables along with their suggested measurement scales.

3.2 Data collection

The collection of sufficient and relevant data is a crucial factor in ensuring the quality of the research and should be compatible with the preferred analysis technique. In this study, a structured questionnaire was designed in two parts to collect information on both qualitative and quantitative measurements. Part A of the questionnaire queries the respondent's experiential knowledge of procurement systems and its sub-systems, while Part B seeks detailed information on a recently completed project in which the respondent participated. After some feedback on the initial questionnaire, its format was modified and dispatched to a random sample covering clients, contractors, consultants, architects, and quantity surveyors in the Hong Kong construction industry.

225 questionnaires were distributed to the random sample, but only 15 responses were received before the deadline. After follow-up with the latter organisations on the telephone, another 5 responses were received, yielding a total of 20 responses from the random sample. Meanwhile 30 questionnaires were distributed among a 'selected sample' (of senior industry personnel who were known to the authors), to which 26 responded. Fourteen project managers (in the selected sample) who agreed to discuss their experiences on this subject, were subsequently interviewed to elicit further details. A total of 46 responses were thus received from both the selected and random samples, but only 32 of them were found suitable to be used for the analysis. The remaining responses were discarded due to missing information and projects being incomplete or if the data did not relate to building projects. The foregoing questionnaire survey was conducted during April-December 96.

Meanwhile, 27 responses (data sets) were used to develop a time over-run model while 25 responses were used to develop a cost over-run model. For testing the models, 5 responses were used for the time over-run model while 3 responses were used for the cost over-run model. The reason for using different responses for the cost and time over-runs was due to the non availability of final accounts of some of the responses at the time of analysis.

3.3 Project performance measures

Time, cost, quality targets and participant satisfaction criteria are the commonly used project performance measures. A project is considered an overall success if it meets the technical performance specification and/ or mission to be performed, and if there is a high level of satisfaction concerning the project performance among principal participants in the project team and key users [De Wit, 1988]. In this study, time and cost over-run indicators are used as first order project performance measures for simplicity.

$$\textit{Time over-run} = \frac{\textit{Actual duration}}{\textit{Programmed duration}} \times 100 \quad \text{equation (1)}$$

$$\text{Cost over-run} = \frac{\textit{Final cost}}{\textit{Original Cost Estimate}} \times 100 \quad \text{quation (2)}$$

3.4 Initial analysis

According to the large number of micro variables and the available sample size (32), neither statistical nor knowledge-based system techniques could be directly applied to analyse the model because of the difficulty of identifying actual correlations between variables. 'Principal Component Analysis' and 'Factor Analysis' have been found to be potentially useful variable reduction techniques [Beck, 1994]. In this study, the Factor Analysis technique was applied but did not yield satisfactory results because of the low correlation between variables, perhaps due to the small sample size. The results are thus not reported in this paper. An alternative method was next adopted: sets of micro variables were identified in relation to each of the selected macro variables. The micro variables corresponding to each macro variable were next separated according to whether they were on an ordinal scale or not.

Finally, all the 'measures' of micro variables on an ordinal 1-5 scale ("5" indicating very high and "1" indicating very low) were combined to form a representative value. This representative value was computed as follows:

First, Spearman's rho correlation coefficients between ordinal scale micro variables and the time (and cost) indices were computed ($\rho_1 \ldots\ldots \rho_n$). The signs of the coefficients were reversed and they were then divided by their maximum absolute value to derive the corresponding weightings to the ordinal scale micro variables ($w_1 = -\rho_1/\rho_{max}, \ldots\ldots\ w_n = -\rho_n/\rho_{max}$). Next, the ordinal scale data of micro variables, corresponding to macro variables (client/ client representative characteristics, contractor characteristics, design team characteristics, and external factors), were consolidated to form a collective representative value for the corresponding macro variable using the following equation.

$$\text{Representative value} = \left[\frac{\sum_{i=1}^{N} W_i X_i}{M \times N}\right] \times 100 \quad \text{equation (3)}$$

where:

X_i = Score of i^{th} Variable

N = Number of Variables

M = Highest Possible Score

W_i = Weighting of i^{th} variable

The project complexity was one of the micro variables of project characteristics but it also represented five variables. Therefore the same formula was used to consolidate those variables.

The micro variables were thus consolidated using the above formula and the remaining residual micro variables (which were measured on nominal and interval scales) are shown in Table 1. However, two micro variables (ie two procurement sub-systems: selection methodology and work packaging) were removed from the analysis in the case of the procurement system macro variable because their variations were found to be insignificant in the database. On the other hand, two micro variables, 'type of construction' and 'gross floor area' were removed from the project characteristics macro variable because the former micro variable had no significant variation, while there were gaps in the data for the latter micro variable.

Table 1:

The selected macro and micro variables for the Multiple Regression Analysis

Macro Variables	Micro Variables
(A) Procurement System	(1) Functional Grouping (2) Payment Method (3) Contract Conditions
(B) Project Characteristics	(4) Building Type (5) Number of Stories (6) Number of Blocks (7) Project Complexity Representative Value* (covers the variables B8-B12 in the Appendix 2) (8) Programmed Duration (9) Original Cost Estimate
(C) Project Team Performance	(10) Contractor (11) Design Team (12) Management Team

Table 1 - Contd.

(D) Client/ Client Representative's Characteristics	(13) Client Type (14) Client Priority (15) Client Source of Finance (16) Client Characteristics Representative Value* (covers the variables D6-D28 in Appendix 2)
(E) Contractor Characteristics (F) Design Team Characteristics	(17) Contractor Characteristics Representative Value* (covers the variables E1-E15 in the Appendix 2) (18) Design Team Characteristics Representative Value* (covers the variables F1-F5 in the Appendix 2)
(G) External Conditions	(19) External Conditions Representative Value* (covers the variables G1-G6 in the appendix 2)

* Different representative values were used for the time and cost over-runs analysis

4.0 Analysis

4.1 Applying Multiple Regression Analysis

Multiple Regression Analysis was considered to be a potentially useful statistical approach to analyse the strength of association between independent and dependent variables; in relation to the main objectives of the research. Therefore it was carried out to regress the 19 independent variables, against each of the first two selected dependent variables- time over-run and cost over-run - using SAS statistical software. One of the most important steps of this regression analysis was to check and validate the model assumptions. The scale of measurements of the independent variables are shown in Appendix 1 and 2 while the representative variables are measured as percentages. The feed forward method was used to deduce the number of insignificant variables from the model. It was started from the highest insignificant variable up to the variable which is significant at level 0.1. This process was repeated until higher explanatory power (R^2 and Adj. R^2) was reached.

One of the problems often associated with Multiple Regression Analysis is the multicollinearity, that may camouflage the results. This is caused when an explanatory variable is virtually a linear combination of other explanatory variables in the model [SAS/INSIGHT User's Guide, 1993]. 'Tolerance'(TOL) and 'variance inflation factor' (VIF) which measure the strength of inter-relationships among the explanatory variables in the model were used to check the multicollinearity. For example, if all the variables are orthogonal to each other, both tolerance and variance inflation are 1 (unity). If a variable is closely related to other variables, the tolerance tends towards zero and the variance inflation becomes large. There was no significant correlation between the independent variables in both the time and cost over-run models. Another model assumption - that the sample distribution is normal, was checked by Kolmogorov's test [SAS/INSIGHT User's Guide, 1993]. Satisfying model assumptions, the time and cost over-run models were found to be significant at 99% confidence interval and the explanation capabilities of models were reasonably good. Therefore the ensuing model equations (as follows) were thus considered sufficiently reliable.

Time Over-Run Equation

R - Square = 0.8493

Adj. R- Square = 0.8040

Root MSE = 10.3188

Sample size = 27

F Sig. = 0.0001

TOR = 77.8875 + 4.0573 PCOMPLEX - 1.0573 PDUR - 21.3222 P1 - 16.0903 P2 - 14.3525P3 + 27.2936 P4

where:

TOR = Time over-run

PDUR = Programmed duration

PCOMPLEX = Project Complexity representative value

CLTYPE = Client Type

- P1 = Property & Development Company
- P2 = Investor
- P3 = Local/ Central Government Authority
- P4 = Occupier

The above individual variables were significant at the following levels: programmed duration ($p<0.0014$), project complexity representative value ($p<0.0001$), and client type ($p<0.0001$).

Cost Over-Run Equation

R - Square	=	0.7886
Adj. R- Square	=	0.7464
Root MSE	=	6.6121
Sample size	=	25
F sig.	=	0.0001
COR	=	71.7383 + 1.4237 PCOPMLEX - 1.4138 CLCHAR - 0.0087 ORICOST + 2.3666 CONCHAR

where:

COR	=	Cost Overrun
PCOMPLEX	=	Project Complexity Representative Value
CLCHAR	=	Client/ Client representative's Char acteristics Representative Value
ORICOST	=	Original Cost Estimate
CONCHAR	=	Contractor Characteristics representative value

The above individual variables were significant at the following levels: project complexity representative value ($p<0.0001$), client/ client representative's characteristics representative value ($p<0.0028$), contractor characteristics representative value ($p<0.0042$) and original cost estimate ($p<0.1066$).

Note: 'Client Characteristics Representative Value' does not represent client type, client priority etc. It only covers the variables (D6-D28) in the Appendix 2.

4.2 Applying Artificial Neural Networks Technology

Various knowledge-based systems have been developed to help decision making and problem solving processes in the construction industry. Artificial Neural Networks (ANN) is another potential approach for somewhat different applications. Generally ANNs are well suited for analysis and prediction of complex pattern or highly erratic data with a large number of independent variables [McKim el at., 1996]. ANNs are analogous to the human brain and the nervous systems. Like the human brain, ANNs learn from experience and generalise from previous examples to a new one if exposed to a new situation. The basic attributes of ANN may be divided into the Architecture and the functional properties or Neurodynamics. Architecture defines the network structure, that is, the number of nodes (Artificial Neurons) and their connectivity. The neurodynamics of neural networks define their properties, that is, how the ANN learns, recalls, associate and continuously compares the new information with existing knowledge. Many successful applications of this technique are found in the field of Civil Engineering (eg. hydro-meteorological modelling [Jayawardena and Fernando, 1995], rainfall-runoff process modelling [Smith and Eli, 1995], soil mechanics [Goh, 1995] and hydraulics [Grubert, 1995]). Examples in the field of construction management are also emerging (eg. mark-up estimation [Li, 1996], predicting project duration [Boussabaine and Cheetham, 1995], predicting variations [Akinsola, at el, 1996], bankruptcy prediction [Khosrowshahi and Elhag, 1996], contractor pre-qualification [Taha, 1994] and organisational effectiveness [McKim et al., 1996]).

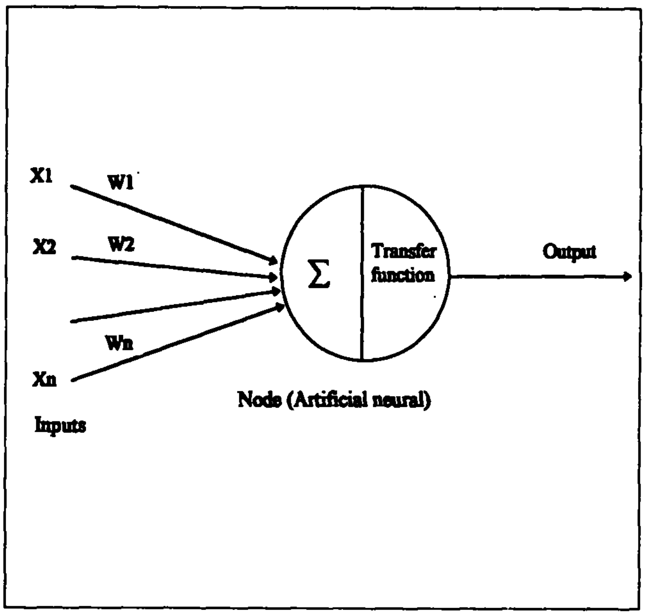

Figure 2: Model of an artificial neurone

The ANN is a parallel information processing system built of many nodes. As shown in Figure 2, the node performs three basic functions. Firstly it receives inputs (X1, X2,...Xn) from other nodes connected with a weight (W1, W2,...Wn) and computes a weighted sum of the inputs and finally the weighted sum passes through the transfer function (mathematical function which acts as non linear threshold) to produce a signal output.

There are different types of ANN available based on their structures and the learning processes. The multilayer neural network with a back propagation learning algorithm, is one such type.

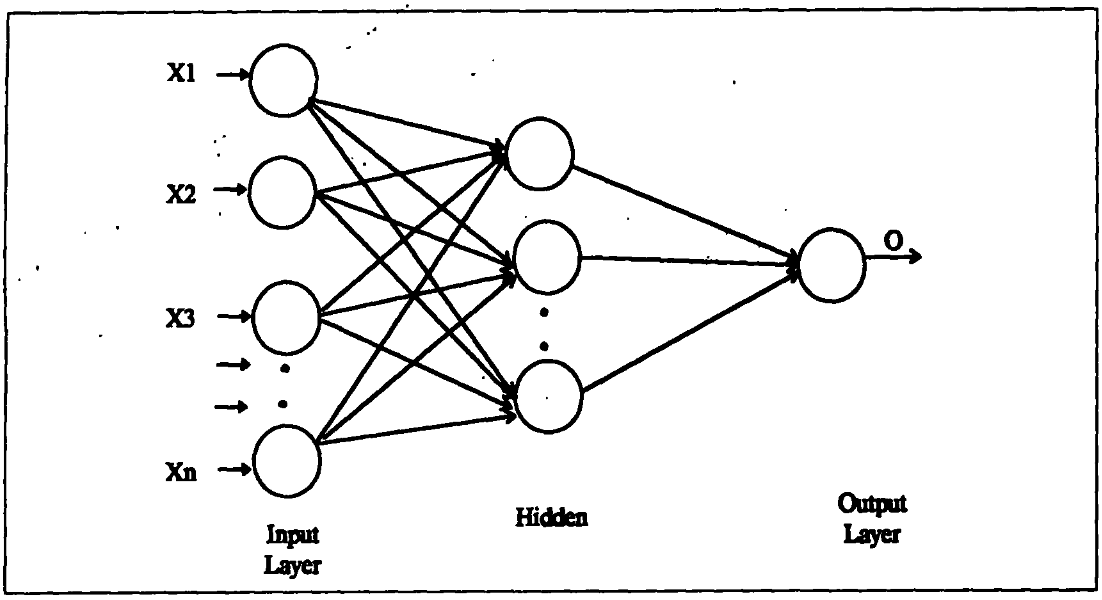

Figure 3: Architecture of a simple Artificial Neural Network

Notes:

O = Output
x = Input variables
m = Number of nodes in hidden layer (varies from 2 to 10)
n = Number of inputs

Figure 3 shows a simple structure of an ANN that represents the type used in this study. It consists of three layers (input, hidden and output) of nodes or artificial neurons. The input variables enter at the input layer, transform at the hidden layer and the outgoing information is produced at the output layer. The nodes in the input layer are connected to those in the hidden layer and they are in turn connected to those in the output layer. The connections are weighted according to their relative strengths.

An ANN has to be trained before it can be used. Training is similar to optimisation of parameters in an ordinary model. The training method used in this study is known as the back propagation of error method. The training process consists of two steps: A feed forward process where the input values are fed to the network and pass through the hidden layer to the output layer to produce an output value. The training algorithm then compares the value produced from the network with the actual value. This is then followed by a feed back process where the error is propagated back via the hidden layer to the input layer. In the process, the weights of the connections are updated. A learning rate determines the step size by which the weights are updated. Training is repeated until the network output error converges to a predetermined level.

From initial modelling exercises, it was found that ANN is not an appropriate prediction method to use with a small size of sample where there is a large number of input variables involved. It appears that there should be a large number of data sets to train the network to a satisfactory level. Therefore in this study, instead of the initial 19 input variables, the significant variables identified from the Multiple Regression Analysis were used as the input variables for the ANNs. The variables shown in Table 2 were used to predict the performance measures of time over-run and cost over-run respectively.

Table 2:
The significant variables selected for Artificial Neural Network Analysis

Performance Measures	Independent Variables
Time over-run	(1) Client type (2) Original programmed duration (3) Project complexity representative value
Cost over-run	(1) Original cost estimate (2) Client characteristic representative value (3) Contractor characteristic representative value (4) Project complexity representative value

The scales of measurement of the above independent variables were as follows: the client type was coded from 1-5 while the project complexity representative value was measured as a percentage and the programmed duration was measured in months. The representative values of project complexity, contractor and client were expressed as percentages. The original cost estimate was measured in millions (HK$). Both project outcome variables: time over-run and cost over-run are continuous and were expressed as percentages.

The selected ANN structures contained only one output node, while the input layers contained 4 nodes for the time over-run, and 3 nodes for the cost over run. Since there is no rule for guiding the best ANN structure for a particular problem, a trial and error method was used to select the number of nodes in the hidden layer. Bailey and Thompson [1990] suggested a rule of thumb to determine the number of nodes for a single hidden layer: The rule suggested that 75% of the number of nodes in the input layer to be adequate for the hidden layer. However in both the exercises, the number of nodes in the hidden layer was varied from 2 to 10 until an optimal solution was reached. The optimum value was reached when there were five nodes in the hidden layer. Training was expected to be terminated either when the total number of training iterations exceeded 2500 or when the error reduced below 0.01. However, the error level attainable in these cases was 0.1, which corresponded to 1500 and 2500 training iterations in the time over-run and cost over-run models, respectively. The learning rate (the rate at which the connection 'weights' are incrementally adjusted) was set to 0.1 for both exercises.

The training of the time over-run model was carried out using 27 out of 32 sample data sets; while 25 out of 28 sample data sets were used for the cost over-run model. As indicated in section 3.2, the lower number of data sets used for the cost over-run exercise, was due to incomplete final cost data on some projects. The testing of trained models were carried out in two stages. In the first stage, the same data sets were used for training as well as testing. In the second stage, five data sets were used for the time over-run, while three were used for the cost over-run for the testing of each of the two performance measures. Tables 3 and 4 show the estimates and predictions by ANN and MR for the training and testing data sets respectively.

Table 3

The estimates of time over-run and cost over-run by the Multiple Regression and Artificial Neural Network methods, using the training data set

Time over-run						Cost over-run					
No	Actual	Prediction		Error		No	Actual	Prediction		Error	
		MR	ANN	MR	ANN			MR	ANN	MR	ANN
1	100.0	89.3	100.0	10.7	0.00	1	103.3	101.7	106.9	1.6	-3.6
2	100.0	104.1	102.3	-4.1	-2.31	2	101.9	93.4	102.1	8.4	-0.2
3	100.0	110.1	100.0	-10.1	0.00	3	101.3	101.1	103.2	0.2	-2.0
4	105.5	107.7	100.0	-2.2	5.50	4	100.0	105.0	100.2	-5.0	-0.2
5	107.1	99.0	100.0	8.1	7.10	5	100.0	107.6	102.5	-7.6	-2.5
6	100.0	103.1	100.0	-3.1	0.0	6	100.0	90.5	95.2	9.5	4.8
7	200.0	192.2	200.0	7.8	-0.0	7	117.1	113.7	116.3	3.4	0.8
8	107.7	118.4	100.1	-10.7	7.6	8	71.4	76.7	72.0	-5.3	-0.6
9	150.0	134.4	150.1	15.7	-0.1	9	90.2	98.5	90.4	-8.3	-0.2
10	100.0	113.5	103.0	-13.5	-3.0	10	130.0	130.1	129.9	-0.1	0.1
11	133.3	121.9	133.1	11.4	0.2	11	100.0	106.5	99.1	-6.5	0.9
12	103.4	114.3	103.3	-10.9	0.1	12	97.1	90.9	99.4	6.2	-2.3
13	120.2	120.9	119.5	-0.7	0.7	13	99.2	102.1	98.0	-2.9	1.3
14	109.1	119.5	109.8	-10.4	-0.7	14	105.8	96.5	107.4	9.3	-1.6
15	100.0	108.1	100.0	-8.1	0.0	15	105.9	100.8	107.6	5.1	-1.7
16	120.0	110.5	119.9	9.5	0.1	16	120.0	117.7	119.0	2.3	1.0
17	114.3	116.6	114.0	-2.3	0.3	17	77.8	85.7	78.1	-8.0	-0.3
18	137.9	137.8	138.2	0.1	-0.3	18	110.0	104.5	109.0	5.5	1.0
19	136.5	144.1	136.4	-7.6	0.1	19	89.1	93.1	89.6	-4.0	-0.5
20	140.0	139.2	139.9	0.8	0.1	20	97.0	88.9	96.1	8.1	0.9
21	103.0	95.6	100.0	7.4	3.0	21	120.0	115.9	120.0	4.1	0.0
22	154.0	154.5	153.7	-0.5	0.3	22	86.2	86.3	88.3	-0.2	-2.1
23	109.5	114.7	109.8	-5.2	-0.3	23	98.8	98.3	98.0	0.5	0.8
24	100.0	112.5	102.6	-12.5	-2.6	24	85.0	93.4	88.2	-8.4	-3.1
25	114.2	100.7	114.2	13.5	-0.0	25	101.2	108.4	101.6	-7.2	-0.4
26	120.0	104.9	100.0	15.1	20.0						
27	100.0	96.9	100.0	3.1	0.0						

Table 4

The prediction of time over-run and cost over-run by the Multiple Regression and Artificial Neural Network methods, using the testing data set

Time over-run						Cost over-run					
No	Actual	Prediction		Error		No	Actual	Prediction		Error	
		MR	ANN	MR	ANN			MR	ANN	MR	ANN
1	106.7	163.7	200.0	-57.0	-93.3	1	87.0	108.9	118.4	21.9	-31.4
2	128.6	79.4	100.3	49.2	28.3	2	136.8	105.2	98.2	31.7	38.6
3	185.0	166.6	200.0	18.4	-15.0	3	98.3	86.5	77.1	11.8	21.2
4	120.0	154.9	200.0	-34.9	-80.0						
5	126.7	157.3	200.0	-30.6	-73.3						

4.3 A comparative study

The accuracies of the predictions by the ANN and MR methods were compared by using three quantitative accuracy measurements: (a) percentage error (PE), (b) Mean absolute percentage error (MAPE) and (c) sum of square errors (SEE). These measurement were calculated as follows:

(1) Percentage error (PE)

$$PE = \left[\frac{P-A}{A} \right] \times 100\,\% \qquad \text{equation (4)}$$

(2) Mean absolute percentage error (MAPE)

$$MAPE = \sum_{i=1}^{n} \left| \frac{PE_i}{n} \right| \qquad \text{equation (5)}$$

3. Sum of the square errors (SEE)

$$SEE = \sum_{i=1}^{n} (P-A)^2 \qquad \text{equation (6)}$$

where:

A = the actual value
P = the predicted value
n = the number of projects considered

Table 5:

Results from the comparative study

	Measures of Accuracy (Error)	Time Over-run		Cost over-run	
		ANN	MR	ANN	MR
Exercise 1	MAPE	1.8	7.9	1.3	5.3
	SEE	569.4	2152.0	79.3	874.4
Exercise 2	MAPE	48.4	31.0	28.6	20.1
	SEE	21500.1	8159.3	2924.3	1620.5

Notes:

'Exercise 1' indicates the training data set
'Exercise 2' indicates the testing data set

5.0 Discussions and Conclusions

The foregoing results suggest that the impact of procurement sub-systems by themselves on project performance, is not as significant as that of other variables which affect project performance. For example, the time over-run equation indicated that project complexity representative value, programmed duration and client type are highly correlated with the time over-run; whilst the cost over-run equation suggested that project complexity representative value, client/ client representative's characteristics representative value and contractor characteristics representative value are highly correlated with the cost over-run.

Previous research had led to parallel observations, although through different approaches. For example: it has been found that the experience of the building team with the building process had considerable impact on the time and cost over-run [Naoum and Musthapa, 1994]. Walker [1994] found that the contract type does not significantly affect speed of construction, while several client related factors and team communications proved more significant. The quality of the relationships between client, client representative, design team and construction management team were thus found to be significant in governing construction time performance. Organisational and contextual/ environmental variables were also found to be strongly correlated with performance [Rowlinsion, 1989].

As shown in Tables 3 and 4, ANN performed better for training (model development). However its prediction capability for the testing data sets seemed much weaker than the prediction capability through the MR exercise. Still this may not yet be considered a firm conclusion due to the somewhat small numbers in the training and testing data sets which were used for the analysis. Nevertheless, it was observed that the ANN procedures performed better than the MR exercises for developing both the time and cost over-run models; while the MR models give better predictions.

The following factors may have possibly contributed to the foregoing outcomes: A simple neural network structure may not be suitable to model a highly complex pattern. The size of the data set used for training is relatively small, compared to the numbers of variables. When the patterns between the independent and dependent variables are highly complex, more data is needed to train the network to recognise the underlying patterns. Even though the network may be well trained to model a given data set, it may still be unable to accommodate a new data set. Another reason for the shortcoming in pattern recognition is perhaps the lower applicability of neural networks when the problem is strongly characterised by particularly variable human/ subjective judgements. Under these constraints the predictability capabilities of ANNs may probably be improved through the use of a much larger training set.

According to the observations from the Multiple Regression Analysis and the Artificial Neural Network modelling, the former provided better results than the latter in time and cost over-run predictions for the testing data sets from the studied sample in Hong Kong. Meanwhile, it was noted that non-procurement related variables (such as project complexity representative value, client type, contractor characteristics representative value, client/ client representative's characteristics representative value, original cost estimate and programmed duration) appeared to exert a more significant influence on project performance than procurement related variables, in terms of time and cost over-runs in the building projects studied in Hong Kong.

However, the sample size that was used for the analysis appears insufficient to regard these observations as firmly conclusive. Therefore the next step in this ongoing research is to focus on further data collection, testing and justifying the use of these two models by including more data to enhance the trainability of the Artificial Neural Networks, for example; as well as to study the identified significant factors in more depth and detail.

6. 0 Acknowledgements

The authors record their appreciation of the valuable assistance rendered by Ms. Achela Fernando who provided the Artificial Neural Network (ANN) 'shell' developed by her, which was used as the basis for the ANN exercises described in this paper.

7.0 References

Akinsola, A. O., Potts, K. F., Ndekugri, I. and Harris, F. C., 1996, "A Neural Network Model for Predicting Building Projects Contingency Allowance", ARCOM 96 Conference, Sheffield, U. K., pp 507-516.

Ashely D. B., Lurie C. S. and Jaselis E. J., 1987, "Determinants of Construction Project Success", Project Management Journal, Vol. 18, No. 2, pp 69-79.

Bailey, D. L. and Thompson, D. M., 1990, "How to Develop Neural Network Applications", AI Experts, Vol. 5, No. 6, pp 38-47.

Beck M. S. L., 1994, "Factor Analysis and Related Techniques, International Handbooks of Quantitative Applications in the Social Sciences", Vol. 5, SAGE Publications, USA.

Boussabaine, A. H. and Cheetham, D. W., 1995, "Artificial Neural Networks as a Tool for Predicting Project Duration", ARCOM 95 Conference, Eleventh Annual Conference, University of York, pp 551-558.

De Wit, A., 1988, "Measurement of project success", Project Management Journal, Vol. 6, No. 3, pp 164-170.

Goh, A. T., 1995, "Neural networks for Evaluating CPT Calibration Chamber Test Data", Microcomputers in Civil Engineering, Vol. 10, pp 147-151.

Grubert, J. P., 1995, "Application of Neural Networks in Stratified Flow Stability Analysis", Journal of Hydraulic Engineering, Vol. 121, No. 7, pp 523-532.

Hashim, M., 1996, "The Effects of Procurement Systems on Performance of Construction Projects in Malaysia", CIB W92, North Meets South, Procurement Systems Symposium, Durban, South Africa, pp 192-199.

Hughes, W. P., 1989, "Identifying the Environments of Construction Projects", Construction Management and Economics, Vol. 7, No. 1, pp 29-40.

Ireland, V., 1985, "The role of managerial actions in the cost, time and quality performance of high-rise commercial building projects", Construction Management and Economics, Vol. 3, No. 1, pp 59-87.

Jayawardena, A. W. and Fernando, A. D. A. K., 1995, "Artificial Neural Networks in Hydro-Meteorological Modelling", Civil-Comp Press, Edinburgh, U.K., pp 115-120.

Khosrowshahi, F. A. and Elhag T., 1996, "Neural Network Model for Bankruptcy Prediction of Constructing Organisations", ARCOM 96 Conference Proceedings, ARCOM, Sheffield, U. K., pp 200-209.

Kumaraswamy, M. M. and Dissanayaka, S. M., 1996, "Procurement by Objectives", Journal of Construction Procurement, Vol. 2, No. 2, pp 38-51.

Li, H., 1996, "Neural Network Models for Intelligent Support of Markup Estimation", Engineering, Construction and Architectural Management, Vol. 3, No. 1 & 2, pp 69-81.

Liu, M. M. A., 1995, "Evaluation of the Outcome of Construction Projects", Ph.D. Thesis, University of Hong Kong.

Masterman, J. W. E., 1992, "An Introduction to Building Procurement Systems", E & F. N. Spon, London.

McKim, R., Adas, A. and Handa, V. K., 1996, "Construction Firm organisational Effectiveness: A Neural Network-Based Prediction Methodology", The Organisation and Management of Construction: Shaping theory and practice, Vol. 3, pp 247-256.

Nahapiet, H. and Nahapiet J., 1985, "A comparison of contractual arrangements for building projects", Construction Management and Economics, Vol. 3, No. 3, pp 217-231.

Naoum, S. G. and Musthapa, F. H., 1994, "Influences of the Client, Designer and Procurement Methods on Project Performance", CIB W92, East Meets West, Procurement Systems Symposium, CIB No 175, Hong Kong, pp 221-228.

Naoum S. G., 1990, "Procurement and project performance", Occasional Paper 45, Chartered Institute of Building, U.K.

Rowlinson, S. M., 1989, "An analysis of factors affecting project performance in industrial building", Ph.D. Thesis, Brunel University, U. K.

SAS/ INSIGHT User's Guide, 1993, Version 6, 2nd edition, SAS institute Inc., U.S.A.

Sharif, A. and Morledge, R., 1996, "Procurement Strategies: The Dependency Linkage", CIB W92, North Meets South, Procurement Systems Symposium, Durban, South Africa, pp 566-577.

Smith, J. and Eli, R .N., 1995, "Neural Network Models of Rainfall-Runoff Process", Journal of Water Resources planning and Management, Vol. 121, No. 6, pp 499-508.

Taha, M. A., 1994, "Applying Distributed Artificial intelligence to the pre-qualification of construction contractors", Ph.D. Thesis, University of Wisconsin-Madison, U. S. A.

Turner, A., 1990, "Building Procurement", Macmillan Education Ltd., London.

Walker, D. H. T., 1994, "An Investigation into factors that determine building construction time performance", Ph.D. Thesis, RMIT, Australia.

Walker, D. H. T., 1995, "The influence of Client and Project Teams relationships upon construction time performance", Journal of Construction Procurement, Vol. 1, No. 1, pp 4-20.

Appendix 1:
List of procurement related variables

Macro variables	Micro variables	Code	Criteria/ measures
	(A.1) Functional grouping	1	Traditional (Sequential)
		2	Traditional (Fast Tract)
		3	Design & Build
		4	Management Contracting
		5	Construction management
Procurement system	(A.2) Payment Method	1	Lump Sum Fixed price (without fluctuations)
		2	Lump Sum Fixed price (with fluctuations
		3	Remeasure
	(A.3) Selection Procedure used for Consultant & contractor	1	Tenders from prequalified parties & lowest bid
		2	Open tenders & Lowest bid
		3	Other means
	(A.4) Contract conditions used	1	HK Government Conditions of Contract
		2	HKIA
		3	Other

Appendix 2:
List of non-procurement related variables

Macro variables	Micro variables	Code	Criteria/ measures
	(B.1) Type of building	1	Residential
		2	Commercial/ Office
		3	Hospital
		4	School
		5	Industrial
		6	Other
Project Characteristics	**(B.2)** Type of construction	1	New
		2	Refurbishment
		3	Extension
	(B.3) Original Cost Estimate		HK$
	(B.4) Programmed Duration		Months
	Size of Project		
	(B.5) No of Stories In numbers		
	(B.6) No of Blocks In numbers		
	(B.7) Total Gross Floor Area		Square feet
	Project complexity **(B.8)** Location, **(B.9)** Design, **(B.10)** Construction, **(B.11)** Coordinative and **(B.12)** Changes (with time)	5	Very high
		4	Moderately high
		3	Average
		2	Slightly low
		1	Very low
Team Performance	**(C.1)** Contractor	5	Very Good
	(C.2) Team, Design Team	4	Good
	(C.3) Management Team	3	Average
		2	poor
		1	Very Poor

	(D.1) Client type	1 2 3 4 5	Developer Investor Local/ Central government Authority Occupier Other (mixed)
	(D.2) Client sources of Finance	1 2 3	Public Private Mixed
Client Characteristics	(D.3) Priority On Time On Cost	1-10	Scale of 1-10 (10 is highest)
	(D.4) Did the client communicate his requirements and priorities clearly at beginning of the project? (D.5) Were the client requirements and priorities realistic in - terms of (a) Time (b) Cost (c) Quality (d) Scope	1 0	Yes No
	(D.6) Client involvement in the project in the following aspects (a) Planning (b) Decision Making (c) Controlling (D.7) Project Manager* capabilities to handle the project (D.8) Project Manager* experience on project of this type (D.9) Project Manager* authority to authorise (a) Time extension (b) Extra costs (c) Modified specification (D.10) Project Manager* ability to make authoritative decisions (D.11) Project Manager* ability to contribute to the construction process (D.12) Speed of decision making (D.13) Client confidence in the design team (D.14) Client confidence in the construction team (D.15) Effectiveness of information flow in (a) Reporting	5 4 3 2 1	Very high High binding / Average ideas Low Very Low

	(b) Changes (variations)		
	(c) Instructions		
	(D.16) `Effectiveness of Communication with		
	(a) Management Team / Construction Team		
	(b) Management Team / Design Team		
	(D.17) Client incentives to accelerate the project		
	(D.18) Project Team motivation and goal orientation		
	(D.19) Frequency of project meetings		
	(D.20) Frequency of schedule adjustments		
	(D.21) Client interest in safety measures		
	(D.22) Level (degree) of Risk allocated to the contractor		
	(D.23) Level (degree) of difficulty of getting interim payments		
	(D.24) Difficulty of getting Payments for additional works		
	(D.25) Possibility of long term relationship (D.26) Level (degree) of the co-operation between the contractor team and the client team (D.27) Level (degree)of the co-operation between the design team and the client team (D.28) Level (degree) of the co-operation between the project management team and the client's main organisation Contractor Characteristics		
	(E.1) Contractor Management Staff number (E.2) Quality of Management and supervision of contractorís staff		
	(E.3) Contractorís top management involvement in the Project		
	(E.4) Availability of Equipment		
	(E.5) Availability of competent workforce (E.6) Response to instructions (E.7) Effective communications		
	(a) within contractorís own team		
	(b) between contractorís team and design/supervision team	5	Very high
	(c) between contractorís team and project management team	4	High

	(E.8) Financial difficulties during the construction	3	Average
	(E.9) Number of subcontractors involved (E.10) Reporting Frequency	2	Low
	(a) Progress	1	Very Low
	(b) Budget		
	(c) Changes and claims		
	(E.11) Contractor team turnover**		
	(E.12) Control systems Effectiveness		
	(a) Budget		
	(b) Resources		
	(E.13) Contractor involvement in general decision making		
	(E.14) Trustworthiness of contractor (E.15) Level of decentralisation of contractor project organisation		
Design Team Characteristics	(F.1) Design team experience		
	(F.2) Frequency of changes in drawings (F.3) Accuracy of detailing in drawings(F.4) Effective Communication of design details to contractor(F.5) Complexity level of drawings	5 4 3 2 1	Very high High Average Low Very Low
External Factors	(G.1) Physical		
	(G.2) Cultural	5	Very high
	(G.3) Social	4	High
	(G.4) Political	3	Average
	(G.5) Legal	2	Low
	(G.6) Economical	1	Very Low
	(G.7) Others (specify)		

Note

* 'Project Manager' is taken here to mean the clientís representative (either in-house or authorised), overseeing both ëdesigní and 'construction'aspects.