Management and Innovation for a Sustainable Built Environment ISBN: 9789052693958
20 – 23 June 2011, Amsterdam, The Netherlands

USE OF ARTIFICIAL INTELLIGENCE TO PREDICT THE ACCURACY OF PRE-TENDER BUILDING COST ESTIMATE

DR. AJIBADE AYODEJI AIBINU
University of Melbourne
Faculty of Architecture Building and Planning,
Parkville, Victoria 3010.
Australia
(Email: aaibinu@unimelb.edu.au)

DHARMA DASSANAYAKE
Charles Stuart University, Australia
International Centre of Water
Wagga Wagga, New South Wales, 2678
Australia
(Email: ddassanayake@csu.edu.au)

THIEN, VUI CHAU
University of Melbourne
Faculty of Architecture Building and Planning,
Parkville Victoria, 3010.
Australia
(Email: ianthienvc@gmail.com)

Abstract

Pre-tender estimates are susceptible to inaccuracies (biases) because they are often prepared within a limited timeframe, and with limited information about project scope. Inaccurate estimation of project uncertainties is the underlying cause of project cost overruns in construction. Typically, cost engineers and quantity surveyors would add contingency reserve to a pretender estimate in order to account for any unforeseen cost that may arise between the date of the estimate and the projected completion date of the project. The traditional 10% rule of thumb for estimating contingency is subjective - based on experience and expert judgment, and are often inadequate. In the research reported in this paper, we propose that learning algorithms trained to use the known characteristic of completed projects could allow quantitative and objective estimation of the inaccuracies in pretender building cost estimates of new projects. The study assumes that the accuracy in the initial estimate (bias) of a completed project is the difference between the actual project completion costs minus the pre-tender cost forecast expressed as a percentage of the actual project completion costs. A three- layer ANN model of feed- forward type with one output node was constructed and trained to generalise nine characteristics of 100 completed projects and the cost data from those projects. The nine input variables of the model are project size (measured by number of storeys and gross floor area), principal structural material, procurement route, project type, location, sector, estimating method, and estimated sum. Estimate accuracy (bias) was used as the output variable. The prediction power stands at 73% correlation coefficient, 3% of Mean Absolute Error and 0.2% Mean Squared Error. It was found that in more than 73% of the test cases the predicted estimate bias did not differ by more than 8.2% from the expected (Maximum Absolute Error). This means that amount of estimate bias predicted by the ANN are similar to what actually occurred. The trained ANN model can be used as a decision *making tool by cost advisors when forecasting building cost at the pretender stage. The model can be queried with the characteristics of a new project in order to quickly predict the error in the estimate of the new project. The predicted error represents the additional contingency reserve that must be set aside for the project in order to cater for possible cost overruns. The model can also be extended to forecast the likely cost of a project.*

Key words: accuracy, Artificial Neural Network , cost, estimating, Modeling, pre-tender,

INTRODUCTION

A large part of the quantity surveyor and cost engineer's role in the construction industry is to provide certainty of cost through the estimate process. Pre-tender cost estimation (or early stage cost estimation) is the forecasting of the cost of a project during the planning and design stage (Serpell, 2005). At the pre-tender stage, project owners are interested in knowing the total project cost commitments. However, cost estimation at the pre-tender stage is susceptible to inaccuracies (biases) because they are often prepared within a limited timeframe, and without finalized project scope. Underestimated pre-tender building cost estimate could lead to a non-viable project being pursued by the project owner or if pursued may lead to project failure. On the other hand overestimate could lead to a viable projects being dropped or re-tendered when there is no bid close enough to permit project award.

One of the methods of increasing the accuracy of a pre-tender estimate is to add an appropriate contingency reserve to the estimate in order to account for any unforeseen costs that may arise between the date of the estimate and projected completion date of the project. There are numerous methods for estimating contingency reserve namely percentage allowance (traditional method), Monte Carlo, artificial neural networks (ANNs), fuzzy logic and regressions (see Baccarini, 2005 for a detail review). The most common method is the traditional 10% rule of thumb (Baccarini, 2005). The approach is subjective - based on experience and expert judgment of the cost engineer. They are arbitrary, difficult to justify or defend (Yeo, 1990) and are often inadequate. The disadvantage of the fuzzy approach is that the relationships between the output (cost estimate) and the inputs (cost drivers) are developed from qualitative information on the project, usually elicited from a knowledgeable expert. Additionally, fuzzy relationships are not primarily empirical models (Smith and Mason, 1996).

Regressions and other parametric forecasting models have limitations in that the underlying relationship between the drivers of cost (input variables) and the cost (output variable) are straight forward and too simplistic when compare to the complexity of the real world relationship between those variables. Regressions are based on cost as a function of the variable that has the most significant effects on that cost (Garza and Rouhana, 1995). Regression assumes that the relationships between cost and the drivers of the cost are linear whereas in construction projects the relationship between the cost of construction and the factors influencing the cost are non-linear and sometimes unknown. Thus, while the outcome regression model is easier to analyze, understand, explained and implement, they may produce a less accurate result since the model is far from the real world (Smith and Mason, 1996).

ANNs are non linear and they eliminate the need to find a good cost estimating relationship that mathematically describes cost as a function of the variables that has the most significant effects on the cost (Kim et al, 2004). Also, ANN can model subtle real word relationship between cost and the cost influencing variables even when the natures of those relationships are unknown. Kim et al (2004) discovered that ANNs are viable and are better approach for estimating construction cost. However, the application of ANNs in construction is a relatively new research area (Kim et al., 2004).

Thus the objectives of this study are:

1. to develop an learning model (Artificial Neural Network) for predicting the accuracy of pre-tender building cost forecasts thereby improve the estimation of project contingency reserve.
2. to offer a new methodology and tool to complement existing methods used for improving the accuracy of pre-tender building cost forecasts.

This study is important because the problem of estimate inaccuracies is reflected by the increasingly large number of projects being completed with cost overrun. If the amount of inaccuracy in a pre-tender estimate can be predicted, cost advisors would be able to develop more appropriate contingency reserve for projects. Project owners can be assured of the costs of their project early in the project development process.

ARTIFICIAL NEURAL NETWORK AND COST ESTIMATION

ANNs are purely data driven models which through training iteratively transition from a random state to a final model (Hasangholipour and Khodayar, 2010). ANN doesn't depend on assumptions about functional form, probability distribution or smoothness (Camargo et al, 2003).

The advantages of using ANN include: It allows the learning from previous project cost estimates and outcomes (actual project completion cost). It can model a complex set of relationship between the dependent variables (i.e. output) and the independent variables (i.e. input variables and in this study the drivers of estimate accuracy. Neural network can also accommodate multicollinearity in the independent variables.

ANNs are data-driven self-adaptive methods in that there are few a priori assumptions about the models for problems under study. They learn from examples and capture subtle functional relationships among the data even if the underlying relationships are unknown or hard to describe. Thus ANNs are well suited for problems whose solutions require knowledge that is difficult to specify but for which there are enough data or observations (Zhang et al, 1998).

ANNs can generalize. After learning the data presented to them (a sample), ANNs can often correctly infer the unseen part of a population even if the sample data contain noisy information. As forecasting is performed via prediction of future behavior (the unseen part) from examples of past behavior, it is an ideal application area for neural networks, at least in principle (Zhang et al, 1998).

At the pretender stage of a project, cost forecasting would depend on limited, noisy and approximate information. At that stage, it is also difficult to understand the underlying cost drivers. Also, the relationship between the cost drivers and the cost outcomes could be significantly nonlinear.

ANN can be used to predict project cost overruns and thereby assist management in developing an appropriate contingency (Chen and Hartman, 2000). Examples of the application of ANNs to predict the level of cost overrun/underrun include: Chen and Hartman (2000) used ANN to predict the final cost of completed oil and gas projects from one organisation using 19 risk factors as the input data. It was found that 75% of the predicted final cost aligned with the actual variance i.e. where the ANN model predicted an overrun/underrun, an overrun/underrun actually occurred. The prediction accuracy of ANN outperformed multiple linear regression. Chau et al (1997) used 8 key project management factors to predict the final cost of construction projects. It was found that more than 70% of the examples did not differ by more than one degree of deviation from the expected. Gunaydin and Dogan (2004) used 8 design parameters to estimate the square metre cost of reinforced concrete structure systems in low-rise residential buildings and found that the ANN provided an average cost estimation accuracy of 93%.The research on the application of ANN to predict cost performance often compares the accuracy of ANN with multiple linear regression and in most cases ANN produce more accurate predictions (e.g. Chen & Hartman, 2000; Sonmez, 2004; Kim at al, 2004; Baccarini (2005)

A single biological neuron is not intelligent. A collection of those neurons is made intelligent by making cooperate actions. Collection as a network creates a pattern of inputs to a neural network and processed as a pattern and results as a pattern. The artificial neuron has been modeled mimicking biological neuron similar way working together to produce remarkable results. So Artificial Neural Networks (ANN), is a mathematical model that was developed based on the phenomenon of error minimization. A processing element in ANN, was arranged as a simple model of biological neuron. ANN learning occurs as given in Equation (i), which simply represents the cost function of a desired (actual) and ANN output.

$$\delta_i^p = d_i^p - y_i^p \text{ ....(i)}$$

where d_i^p is the desired i^{th} output of the pattern p and y_i^p is the network's i^{th} output received from the neuron system. The error for all patterns in equation (i) is measured using mean square error (MSE), which is a major performance measurement in ANN learning shown in equation (ii). Lower MSE indicates higher learning of the set of input pattern.

$$MSE = \frac{\sum_p \sum_i \left(d_i^p - y_i^p\right)^2}{N * P} \text{ ....(ii)}$$

where N is number of input data sets and P is number of processing elements.

The learning (training) of ANN model from given inputs and outputs occurs through the iterations. Equation iii shows the network output y of an ANN calculated from n elements of an input pattern x through a summation of weighted inputs and a transfer function.

$$y = F\left(\sum_{i=1}^{n} w_i x_i + \theta\right) \text{ ....iii)}$$

Where x_i denotes i^{th} element of the input pattern x, w_i is the weight for the input x_i, θ is the offset and F is a transfer function, which is a smoothing function.

RESEARCH METHOD

Development of Learning ANN Model

Six steps were followed in the development of the learning ANN model for predicting accuracy of estimates namely:

(1) Defining the ANN output variable – estimate accuracy (or bias)

(2) Identifying the ANN input variables – the characteristics of a building project that could influence the accuracy of pre-tender building cost estimates
(3) Data collection on completed building projects
(4) Development of the learning ANN model
(5) Testing the ANN model: i.e. predicting the accuracy of estimates using new data set.
(6) Evaluating the performance of the learning model and sensitivity analysis.

Defining the output variable – Estimate accuracy (bias)

Skitmore (1991) describes the accuracy of early stage estimate as comprising two aspects, namely, bias and consistency of the estimate when compared with the contract or accepted tender price. Bias is concerned with "the average of differences between actual tender price and forecast" while consistency of estimates is concerned with "the degree of variation around the average". Aibinu and Pasco (2008) examined bias of estimate in terms of the difference between pre-tender forecast and the accepted tender sum expressed as a percentage of the accepted tender sum. This did not account for uncertainties associated with the construction process. Thus it has limited application when attempting to prevent project cost overrun in that upon project award, issue may arise during the construction phase which could result in project cost overrun. Thus Aibinu and Pasco's (2008) approach has limited application that could help ensure project success in the area of cost forecasting and cost performance. To address this problem, this study defined estimate bias as the difference between pre-tender forecast and actual completion cost expressed as a percentage of the actual completion cost.

Thus Percentage cost overestimate or underestimate (estimate error or bias) were estimated by using the following expression:

$$\text{Estimate bias} = \frac{\textit{pretender} \cos t \ \textit{estimate} - \textit{project completion} \cos t}{\textit{project completion} \cos t} \quad x \quad 100$$

A positive value of estimate bias implies an overestimation of cost while a negative value implies an underestimation of cost.

Selection of Input variables: Factors affecting the accuracy of pre-tender building cost estimates

According to Seo et al (2002) attributes used as inputs for ANN model may be derived from the literature. They must be meaningful to the estimator and the design team at the pre-tender stage hence should consist of attributes that are known during that stage. It is also useful that the attributes should, as much as possible, be high-level project characteristics. An overview of previous studies suggests that a large number of variables may contribute to the accuracy of pre-tender cost estimate. Gunner and Skitmore (1999a) reviewed previous studies and summarised the factors as follows: building function, type of contract, conditions of contract, contract sum, price intensity, contract period, number of bidders, good/bad years, procurement basis, project sector (public, private or joint), number of priced items and number of drawings. Gunner and Skitmore (1999a) analysed the estimates of 181 projects in Singapore. They found that a majority of the factors influenced the accuracy of estimates. Using data from 42 projects in Singapore Ling and Boo (2001) found similar results when they compared five variables against Gunner and Skitmore's (1999a) work. Skitmore and Picken (2000) studied the effect that four independent factors (building type, project size, project sector and year) had on estimating accuracy. They tested the four factors using data from 217 projects in the United States of America. They found that bias in the estimate of the projects is influenced by project size and year, while consistency in the estimates is influenced by project type, size and year. In a study of 67 process industry construction projects around the world, Trost and Oberlender (2003) identified 45 factors contributing to the accuracy of early stage estimates. They summarized the factors into 11 orthogonal elements. Of the 11 factors, the five most important include: process design, team experience and cost information, time allowed to prepare estimates, site requirements, and bidding and labour climate. All these studies suggest that there are a large number of variables that may substantially influence the accuracy of an early stage estimate.

According to Gunner (1997) the factors influencing accuracy of estimates are inter correlated so that the true bias of one factor could be masked by one or more factors. For example, Gunner and Skitmore (1999b) theorise that “Price Intensity alone is both necessary and sufficient to account for systematic bias (inaccuracy) in building price forecasting”. Price intensity is the total cost of a building divided by the gross floor area. Price intensity theory states that buildings with low unit rates (cost/m^2 gross floor area) would tend to be overestimated, while those with high unit rates would tend to be underestimated. In a study of 89 construction projects in Hong Kong, Skitmore and Drew (2003) support the price intensity theory. In another study, Skitmore and Picken (2000) using data from 217 projects in the United States found that ‘year’ was the underlying variable responsible for the bias and inconsistency in cost estimates, after partialling out confounding effects of the four factors put forward. The finding contrasts Gunner and Skitmore’s (1999b) ‘price intensity’ theory. However, their result supports Gunner’s (1997) theory which states that intercorrelations among variables cause confounding effects. It also supports Gunner and Skitmore (1999a) in their suggestion that a single underlying variable is the cause of bias and consistency seen in estimates.

In their study of 56 projects in Australia, Aibinu and Pasco (2009) found no evidence to support the price intensity theory. The study defined bias as cost forecasts compared with accepted tender sum. Based on regression analysis, the study discovered that the project size is the only factor that significantly influenced the accuracy of the estimates (bias) of the projects studied. However, project size explained only 29% of the changes in estimate bias with 71% bias unaccounted for. Thien (2008) conducted a regression analysis of the factors that influenced estimate bias on 100 projects in Malaysia. The study defined bias as the cost forecasts compared with final completion cost. Thien (2008) work shows that 7% of the bias in the estimates of the projects can be explained by project type with commercial project tending to significantly suffer most from estimate bias. The study however has 93% of bias unaccounted for.

It appears that in Aibinu and Pasco (2008) and Thien (2008) the regression model was unable to detect the subtle and non-linear relationship between the factors influencing bias and the bias observed; and may have been responsible for the low explanatory power of their models. The use of ANN should produce a better predictive model. In this study, rather than selecting the significant factors discovered by previous studies, we use nine high level project attributes (Table 1) identified from the literature as input variable in the training of the learning ANN model.

Table 1 Project characteristics used in training the learning model (input variables)

Project Characteristics	**Unit**	**Type of information**	**Descriptors**
Gross Floor Area (GFA)	m^2	Quantitative	n.a
Principal structural material	No unit	Categorical	1- steel 2 - concrete
Procurement route	No unit	Categorical	1- traditional 2- design and construct
Type of work	No unit	Categorical	1- residential 2 - commercial 3 - office
Location	No unit	Categorical	1 - central business district 2 - metropolitan 3- regional
Sector	No unit	Categorical	1 - private sector 2- public sector
Estimating method	No unit	Categorical	1- superficial method 2 - approximate quantities
Number of storey	No unit	Categorical	1 – one to two storey(s) 2- three to seven storeys 3 - eight storeys and above
Estimated Sum	Cost/ m^2	Quantitative	n.a

Data collection and processing

Data from 100 construction projects completed over ten years (1999 – 2007) were collected from the office of a quantity surveying firm in Malaysia. Information obtained in respect of each project include: project size (pre-tender cost estimate, final completion cost, the number of floors, and gross floor area - GFA) and other information relating to the input variables (drivers of estimate accuracy) – see Table 1. The 100 projects were identified by a systematic sampling process. Projects that were not suitable for analysis because they were discontinued at the feasibility stage were discarded. The researchers had first hand access to all data, such as estimating report, tender evaluation report, and final accounts. Table 2 shows the profile of the 100 projects data set obtained including the mean estimate bias, standard deviation and the coefficient of variation. The standard deviation and the coefficient of variation (CV) were determined for projects in the different categories of the 9 factors used in the model training. Coefficient of variation is a measure of predictability of estimate bias. Large coefficient of variation implies that estimate bias is volatile and unpredictable. Standard deviation (S) was computed for the projects using the expression:

$$S = \sqrt{\Sigma(x-\bar{x})^2 / n}$$

Where: x = estimate bias; $\bar{x}$ = mean estimate bias
n = number of projects

Thereafter, the consistency in the estimates was determined by calculating the coefficient of variation (CV) as follows:

$$CV = \frac{standard\ deviation}{mean\ estimate\ error} \quad x \quad 100$$

By visual inspection of Table 2, the estimates for the 100 projects are somewhat inconsistent with coefficient of variation ranging from 11.77% to 21.17% across the various project categories. It is assumed that a double digit coefficient of variation is large. Thus the risk of estimation bias is not small. It also suggests that firms have little control over the propensity that estimates would be biased.

Table 2: *Profile of projects according to factor category*

Project factors	Number of Projects	Mean Error (%) (Estimate Bias)	Standard Deviation (%)	Coefficient of Variation(%) (Estimate Consistency)
Estimated Sum (RM)				
1 - 5,000,000	27	2.41%	20.46%	19.98%
5,000,001 - 10,000,000	8	0.63%	18.15%	18.04%
10,000,000	65	0.51%	13.77%	13.70%
Gross Floor Area (m^2)				
1 - 3,000	27	2.33%	21.66%	21.17%
3,001 - 10,000	18	- 4.83%	12.99%	13.65%
above 10,000	55	2.31%	13.39%	13.09%
Number of Storeys				
1-2 storeys	48	1.52%	16.84%	16.59%
3-7 storeys	26	3.58%	17.99%	17.37%
8+ storeys	26	-1.59%	13.35%	13.57%
Location				
CBD	39	-0.31%	18.70%	18.76%
Metropolitan	49	1.55%	14.83%	14.60%
Regional	12	3.25%	12.15%	11.77%
Procurement Route				
Traditional (Lump Sum)	59	0.53%	16.62%	16.53%
Design & Construct	41	1.76%	15.61%	15.34%
Principal Structural Material				
Steel	39	1.15%	16.84%	16.65%
Concrete	61	0.93%	15.82%	15.67%
Estimating Method				
Superficial	64	0.20%	16.11%	16.08%
Approximate Quantities	36	2.47%	16.33%	15.94%
Sectors				
Private	59	0.97%	15.10%	14.95%
Public	41	1.12%	17.72%	17.52%
Project Type				
Residential	50	3.94%	12.57%	12.09%
Commercial	32	-5.63%	19.34%	20.49%
Office	18	4.78%	15.58%	14.87%

The ANN Architecture and Training

Using 85 out of the 100 data sets, different Networks from different topologies of ANN model were trained and optimized, by changing the network parameters in order to obtain a good relationship with better correlation coefficients and performance. The ANN models that were trained include Generalized Feed Forward (GFF), Jordan and Elman Networks, Radial Basis Function (RBF) and Self Organizing Feature Maps.

Among the abovementioned networks, Feed Forward RBF was found to be the best topology that provided significant learning for the aforementioned relationships (Figure 1). The RBF has been constructed using the mathematical function (Neurosolutions 4.32, Lefebvre *et al.*, 2003) in a hidden layer with appropriate number of processing elements. Equation (iv) shows the mathematical form of RBF.

$$G(x;x_i) = \exp\left[\frac{-1}{2\sigma_i^2}\sum_{k=1}^{p}(x_k - x_{ik})^2\right] \qquad \text{(iv)}$$

where G is multivariate Gaussian function, σi is variance of p data points, xi is mean at ith node. Figure (1) is a schematic diagram of the RBF network, shows input layer with two inputs, RBF layer, four hidden layers and output layer with single output.

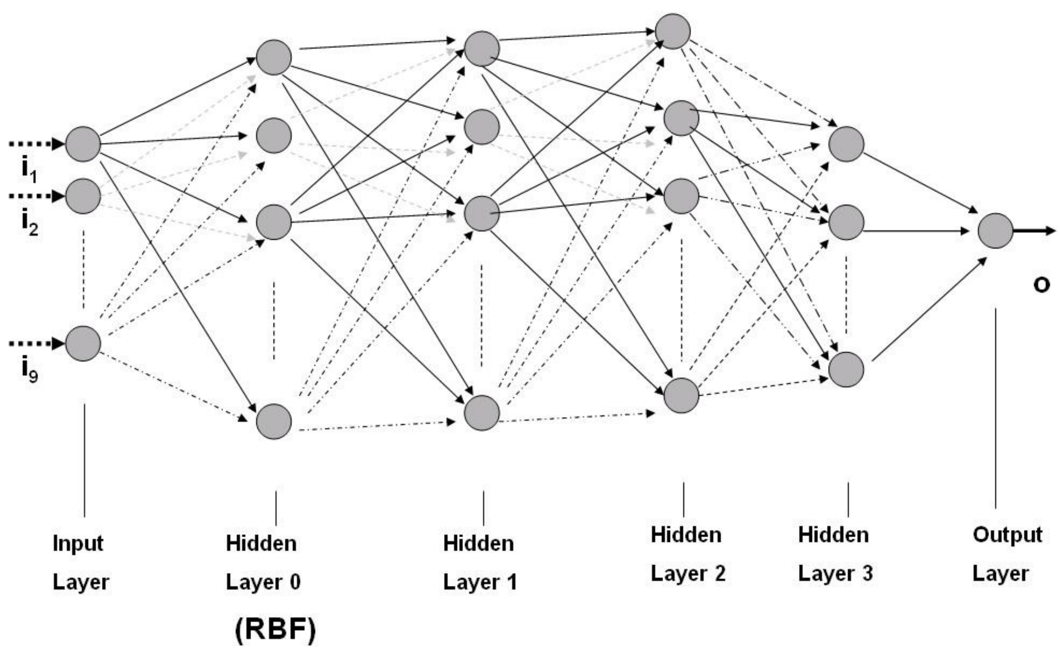


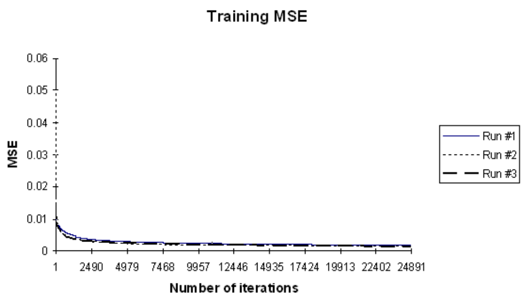
Figure 1: *RBF network trained – Four hidden layers with 100, 80, 20 and 4 PEs in each layer respectively. Hidden layer 0 is RBF.*

The network found has been consisted of four hidden layers including RBF layer and GFF layers. No of processing elements were 100, 80, 20 and 4 for 9 inputs that produced 'estimate bias' single output. Gaussian transfer function was used in the aforementioned RBF layer Tanh transfer functions were used in all other GFF layers.

Input data was normalized and arranged in three sets training, cross validation and testing. The training dataset was used to train networks whilst cross validation data set is evaluating the training. Once a network is learnt, the test data set was used to forecast 'estimate bias' from the network and comparison could be made against actual. Normalized data set is useful in training as each parameter is mapped into a radius of 1 in order to setup a unique boundary.

Results

In this study, out of more than 60 different Networks that were trained and optimized using 85 training and cross validation data sets, Figure 2 and Table 3 shows the performance of the best Network found.



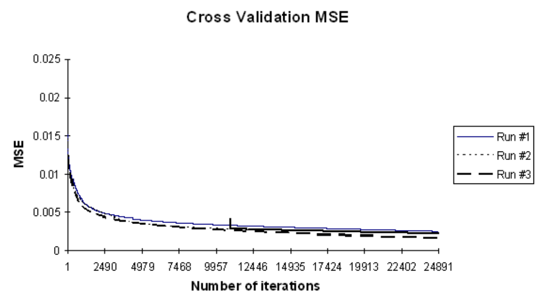


Figure 2: *Performance of the best network*

Table 3: *Performance of the best network*

All Runs	Training Minimum	Training Standard Deviation	Cross Validation Minimum	Cross Validation Standard Deviation
Average of Minimum MSEs	0.0015	0.0002	0.0021	0.0004
Average of Final MSEs	0.0015	0.0002	0.0021	0.0004

Table 4: *Network training and validation*

Best Networks	Training	Cross Validation
Run #	3	3
Epoch #	24899	24899
Minimum MSE	0.0013	0.0017
Final MSE	0.0013	0.0017

Figure 3 shows the actual and predicated values of Estimate bias versus estimated sum for the entire data set. Similarly Figure 4 shows the actual and predicated values of estimate bias versus gross floor area. The correlation coefficient stands at 90%.

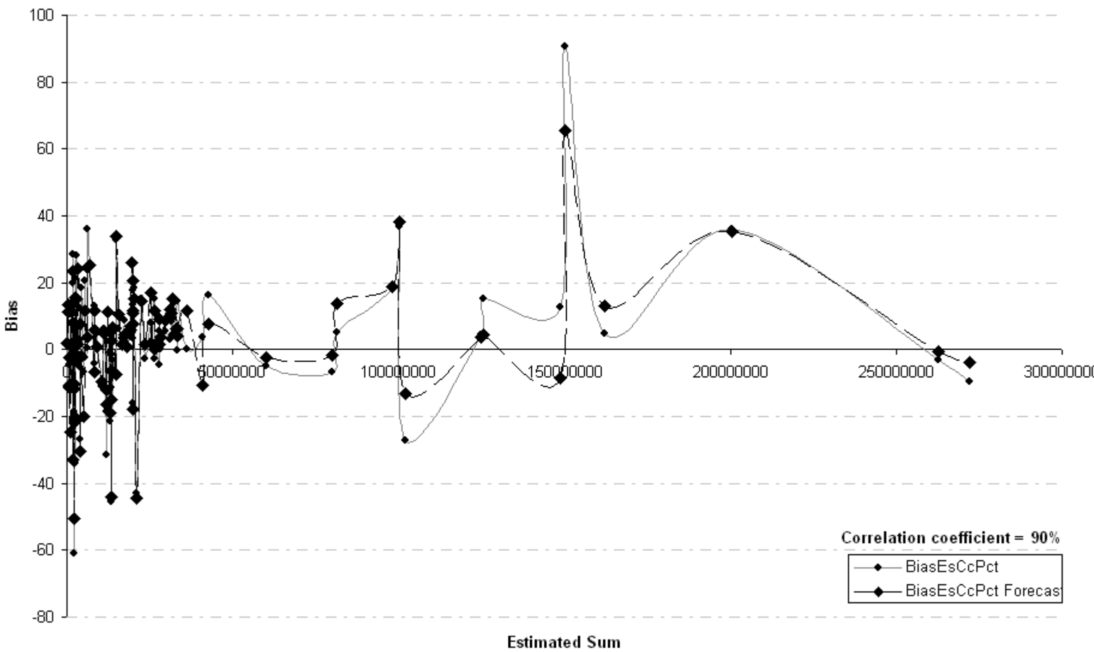


Figure 3: *Comparison between Estimate Bias, actual and network forecast vs Estimated Sum – 90% correlation coefficient*

Model Validation and Performance of ANN

Figure 4 shows the trained network forecast progressively on test data set. The major performance measure used in the training was Mean Squared Error (MSE) which is 0.002 (Table 5). Based on the normalized input dataset the correlation coefficient was found to be 73% while the maximum absolute error is 0.10 (Table 5). This means that in 73% of the test cases, the predicted estimate bias did not differ by more than 10% from the expected. Based on this performance measures the trained network is suitable for forecasting estimate bias and can be extended to forecast actual cost when estimated cost is given.

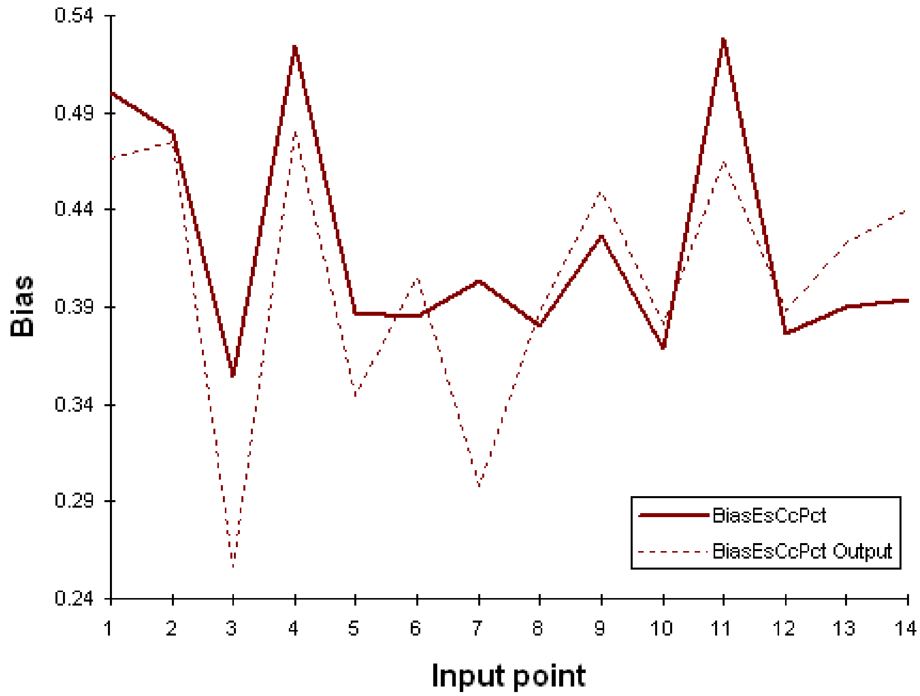


Figure 4: *Network forecast comparison – Estimate bias actual vs. Network prediction*

Table 5: *Model performance measures*

Performance measure	*Value*
Mean Square Error	0.002
Mean Absolute Error	0.039
Min Absolute Error	0.005
Max Absolute Error	0.105
Correlation coefficient	0.73

Contribution of Input Parameters

Some input variables are more effective than others in the estimate. Analysis in this regard is shown in Figure 6 and percentages of contributions are tabulated in Table 6 arranged from highest to the lowest.

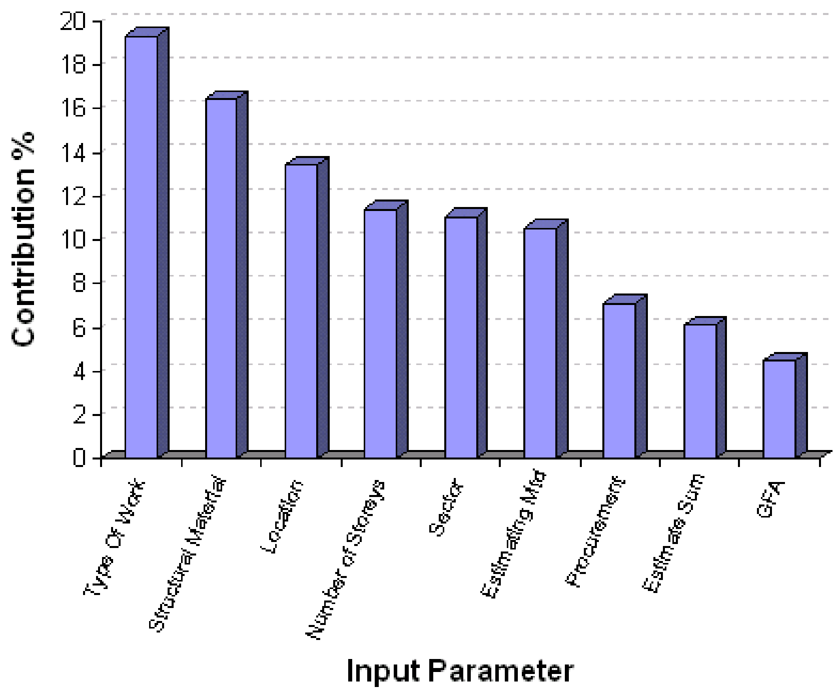


Figure 5: *Input contribution chart*

Table 6: *Input contribution table in percentage*

Input Parameter	Contribution %
Type Of Work	19.31752719
Structural Material	16.46834767
Location	13.42777368
Number of Storeys	11.41395504
Sector	11.08260286
Estimating Mtd	10.54325252
Procurement	7.129983815
Estimate Sum	6.122432462
GFA	4.494124758

Table 6 shows that ‘Type of work’ contributed the highest to estate accuracy whilst ‘GFA’ is the lowest.

Conclusion

At the pretender stage, little information is often available about projects thereby making estimates of contingency allowance difficult. If quantity surveyors can accurately predict the inaccuracies in their estimate they can include contingency allowance to cover such inaccuracies. Learning model can facilitate the prediction of estimate inaccuracies because the modeling assumptions are less rigid when compared to regression modeling. Using ANN models can offer an efficient method of predicting cost at the pretender stage. This study applied neutral network approach for estimating the accuracies in pretender building cost estimates. The trained ANN model can be used as a decision making tool by cost advisors when forecasting building cost at the pretender stage. The model can be queried with the characteristics of a new project in order to quickly predict the error in the estimate of the project. The predicted error represents the additional contingency reserve that must be set aside for the project to cater for cost overruns. The model can also be extended to forecast actual cost of a project when the estimated cost is known. Further study is needed with larger sample size to improve on the prediction power of the model. Future study will also seek the integration of artificial intelligence and probability estimating. Rather than obtain point estimate, the probability estimating model with artificial intelligent will produce a minimum estimate, most likely estimate and a worst scenario estimate of contingency allowance. This should provide more robust information for decision making at the design stage.

References

Hasangholipour, T. and Khodayar, F. (2010) A Novel Optimized Neural Network Model for Cost Estimation Using Algorithm . *Journal of Applied Science*. Vol 10, Issue 6 , pp. 512 – 516.

Camargo, M.B., Rabenasolo, J.M., Castelain, J.M, and Jolly-Desodt, A.M. (2003) Application of the parametric cost estimation in the textile supply chain. *Journal of Textile Apparel Technology Management.* Vol 3, pp 1-12.

Aibinu, A.A and Jagboro, G.O. (2002) The effects of construction delays on project delivery in Nigerian construction industry. *International journal of Project management*, 20, 593 – 599.

Aibinu, A.A. and Pasco, T. (2008) The Accuracy of Pre-tender Building Cost Estimate in Australia. *Construction Management and Economics*, Vol. 26, 12, pp. 1257 – 1269.

Khan, S., Dassanayake, D., and Rana, T. (2005). Ocean based water allocation forecasts using an artificial intelligence approach. in: *MODSIM 2005 International Congress on Modelling and Simulation Advances and Applications for Management and Decision Making, 12-15 December 2005, Melbourne: Modelling and Simulation Society of Australia and New Zealand*: 1667-1673.
http://www.mssanz.org.au/modsim05/papers.htm

Khan, S., Rana, T., Dassanayake, D., Abbas, A., Blackwell, J., Akbar, S., and Gabriel, H. (2008). Spatially distributed assessment of channel seepage using geophysics and artificial intelligence. Irrigation and Drainage, (Published online 13 Aug 2008. doi: 10.1002/ird.415):

[np].
http://www3.interscience.wiley.com/journal/121377494/abstract?CRETRY=1&SRETRY=0

Khan, S., Robinson, D., Beddek, R., Wang, B., Dassanayake, D., and Rana, T. (2004). Hydro-climatic and economic evaluation of seasonal climate forecasts for risk based irrigation management.CSIRO Land and Water Technical Report 5/045/04. Griffith: CSIRO Land and Water. http://www.clw.csiro.au/publications/technical2004/tr5-04.pdf

Zhang, G. Eddy, B. and Hu Y.M. (1998) Forecasting with artificial neural networks:: The state of the art, *International Journal of Forecasting,* Vol. 14, 1, pp.35 – 62.

Baccarini, D. 2005. Estimating Project Cost Contingency - Beyond the 10% syndrome. *Australian Institute of Project Management National Conference, 09/11/2005*. Victoria: Australian Institute of Project Management.

Garza, J. and Rouhana K. (1995) Neural network versus parameter-based application. *Cost Engineering*, 37,25, pp. 14-18.

Kim, G.H.., An, S., Kang, K (2004) Comparison of Construction Cost Estimating Models Based on Regression Analysis, Neural Networks, and Case-Based Reasoning. *Journal of Building and Environment*. Vol. 39, pp 1235-1242.

Baccarini, D. (2004) Accuracy in Estimating project cost construction contingency – a statistical analysis. *RICS COBRA*. 7-8 September Leeds metropolitan University, UK .

AIQS (2003), *AIQS Website: www.aiqs.com.au*. Canberra.

Akintoye, A. and Fitzgerald, E. (2000). "A survey of current cost estimating practices in the UK." *Construction Management and Economics* 18: 161 - 172.

Bromilow, F.J., Hinds, M.F. and Moody, N.F. (1988). *The Time and Cost Performance of Building Contracts 1976-1986*. The Australia Institute of Quantity Surveyors.

Brown, M.B. and Forsythe, A.B. (1974). *Journal of the American Statistical Association*, 69, 364–367.

Cheong, P.F. (1991) Accuracy of design stage cost estimating, MSc Dissertation, National University of Singapore.

Conover, W. J., Johnson, M. E., et al (1981). "A Comparative Study of Tests for Homogeneity of Variances, with Applications to the Outer Continental Shelf Bidding Data." *Technometrics* 23(4): 351 - 361.

Fleiss, J. L. (1971). "Measuring nominal scale agreement among many raters." *Psychological Bulletin* 76(5): 378 - 382.

Gunner, J. (1997). Accuracy of construction pre-bid forecasts, PhD thesis, University of Salford.

Gunner, J. and Skitmore, M. (1999a). "Comparative analysis of pre-bid forecasting of building prices based on Singapore data." *Construction Management and Economics* 17: 635 - 646.

Gunner, J. and Skitmore, M. (1999b). "Pre-bid building price forecasting accuracy: price intensity theory." *Engineering, Construction and Architechural Management* 6(3): 267 - 275.

Kometa, S. T., Olomolaiye, P. O., et al (1994). "Attributes of UK construction clients influencing project consultants' performance." *Construction Management & Economics* 12(5): 433 – 443.

Landis, J. and Koch, G. (1977). "The measurement of observer agreement for categorical data." *Biometrics* 33(1): 159 - 174.

Levine, D.M., Stephan, D. Krehbiel, T.C., and Berenson, M.L. (2005) Statistics for Managers, 4th edition. Prentice hall: New Jersey. Page 429.

Ling, Y. Y. and Boo, J. H. S. (2001). "Improving the accuracy of approximate estimates of building estimates." *Building Research & Information* 29(4): 312 - 318.

Lowe, D. and Skitmore, R. M. (2001). "Human effects in construction contract price forecasting: experience and experimental learning styles." *Journal of Construction Engineering and Management* 127(6): 485 - 493.

McCaffer, R. (1976) Contractor's bidding behaviour and tender price prediction, PhD thesis, Loughborough University of Technology, UK.

Mills, A. (1997) Analysis of the accuracy of building price forecasts. *The Australian Institute of Quantity Surveyors,* 1 (2), 31-38.

Morrison, N. (1984) The accuracy of quantity surveyors' cost estimating. *Construction Management and Economics*, 2, 57 - 75.

Professions Australia’s Education Committee (2005) Skill Mapping: Assessing Australia's Longer Term Requirements for Professional Skills. A Discussion paper.

Rawlinsons (2007), *Australian Construction Handbook 2007*. Perth

Rawlinsons (2006), *Australian Construction Handbook 2006*. Perth

Serpell, A. F. (2005). "Improving Conceptual Cost Estimating Performance." *AACE International Transactions,* EST.13: EST.13.1 - 13.6.

Skitmore, M. (1991), *Early Stage Construction Price Forecasting*. London.

Skitmore, M. and Drew, D. (2003). "The analysis of pre-tender building price forecasting performance: a case study." *Engineering, Construction and Architechural Management* 10(1): 36 - 42.

Skitmore, R. M. and Picken, D. (2000). "The accuracy of pre-tender building price forecasts: an analysis of USA data." *Australian Institute of Quantity Surveyors Refereed Journal* 4(1): 33 - 39.

Smith, P. (1994) Trends in the Australian Quantity Surveying Profession. *A paper presented at the 4th World Congress on Cost Engineering, Project Management and Quantity Surveying, and 18th International Cost Engineering Congress, 17-21 April, 2004, Cape Town, South Africa,* by International Cost Engineering Council.

Thien V.C (2008) Improving the accuracy of preliminary building cost estimate in Malaysia. *A dissertation submitted to the Faculty of Architecture Building and Planning, the University of Melbourne*.

The Australians Department of Employment and Workplace Relations (DEWR) Skilled Occupation Lists (2008) available at http://www.immi.gov.au/allforms/pdf/1121i.pdf. and accessed on 30th May 2008.

Trost, S. M. and Oberlender, G. D. (2003). "Predicting Accuracy of Early Cost Estimates Using Factor Analysis and Multivariate Analysis." *Journal of Construction Engineering and Management* 129(2): 198 - 204.