

Estimation of life-cycle costs of buildings: regression vs artificial neural network.

Ayedh Alqahtani

Department of Civil Engineering, Curtin University, Perth, Australia

Andrew Whyte

Department of Civil Engineering, Curtin University, Perth, Australia

Abstract

Purpose – The purpose of this study is to compare the performance of regression and artificial-neural-networks (ANNs) methods to estimate the running-cost of building-projects towards improved accuracy.

Design/methodology/approach – A data-set of 20 building-projects is used to test the performance of these two (ANNs/regression) models in estimating running cost. The concept of cost-significant-items is identified as important in assisting estimation. In addition, a stepwise technique is used to eliminate insignificant factors in regression-modelling. A connection-weight method is applied to determine the importance of cost factors in the performance of ANNs.

Findings– The results illustrate that the value of the coefficient of determination = 99.75% for ANNs model(s), with a value of 98.1% utilising multiple-regression-model(s); secondly, the mean percentage error for ANNs at a testing stage is 0.179, which is less than that of the mean percentage error gained through multiple-regression-modelling of 1.28; and thirdly, the average accuracy is 99% for ANNs model(s) and 97% for multiple-regression-model(s). On the basis of these results, it is concluded that an ANNs model is superior to a multiple-regression-model when predicting running-cost of building projects.

Research limitations/implications – A means for continuous improvement for the performance of the models accuracy has been established; this may be further enhanced by future extended sample.

Originality/value – This work extends the knowledge base of life-cycle estimation where ANNs method has been found to reduce preparation time consumed and increasing accuracy improvement of the cost estimation.

Keywords Artificial-neural-networks (ANNs), Cost-Significant-Items (CSIs), multiple-regression, A connection weight (CW), Life-cycle estimation, Building projects.

Paper type – Research paper

Introduction

The principle of Life Cycle Costing (LCC) is not new; the Japanese used LCC concepts widely post World-War-II to rejuvenate their economy through specification choices that would last longer and thus save costs in the long term. Implicit within LCC is the need to address a built asset's running costs with reference to ongoing maintenance and operation expenditure. However, the main historic, perceived weakness of this *term* model was its incapability to *predict accurately* or confidently future costs (Boussabaine and Kirkham 2007). Term prediction capabilities increasingly dominated LCC methodologies to assess building costs throughout the late 1970s such that, as a result of increased coverage in the civil engineering sector, it is now widely acknowledged that the initial capital cost of construction represents only a small amount (25-40%) of building life cycle costs.

In order to successfully complete projects and make long term profit, stakeholders now recognise that the acquisition decisions of construction project elements, at the design stage, should be based on life cycle costs rather on initial costs. Appropriate cost reduction measures can be easily taken to predict life cycle cost at an early design phase; however, when the construction project moves from early design stage to construction stage, possibilities to influence the whole construction project cost are decreased quite significantly. The significance of an implementation of a LCC approach is enhanced by injecting the maximum information into the design phase, assisting to decrease waste and to improve efficiency of design and construction as well as operation and maintenance. Recently LCC analysis has, once again, come to the fore as a result of Building Information Modelling (BIM) integrative initiatives that seek to extend design appreciation towards 4D (representing the four

dimensions of width, length, height and time sequencing in a project) and then on into 5D and 6D (life-cycle cost estimating and facilities-management respectively).

All construction-projects are unique with costs dependent upon an estimation of resources usage and specification choice. Risk and uncertainty face design-teams throughout all stages of a building-project's life-cycle; correct acknowledgement of all factors must be taken into account to avoid delay and cost-blowout. Cost-estimation inaccuracy is a major hurdle, where a lack of applied cost data can lead to inadequately completed projects, and failure to realise building objectives (Alqahtani and Whyte 2013). Therefore accurate, reliable cost-predictions should be developed to ensure that projects satisfy the design brief, enhance on-site construction and manage uncertainty throughout an assets' life (Yaman and Tas 2007). The significance and influence of construction cost estimating overall is supported by research: Carr (1989) contends that cost-estimation delivers substantial information for planning, resource controlling and decision making; cost-estimation is deemed key to the success of a project (Dysert and Elliott 2002); Alcabes (1988) articulated that cost-estimators' duties include preparation of all estimates, check-lists and pricing information, asserting that cost-estimation is at the heart of construction work as it involves cost reporting, accurate cost classification and calculation of profit; and indeed, Akintoye and Fitzgerald (2000), argue cost-estimation as a key facilitator for construction contractors for planning purposes.

Markets will increasingly request buildings with low running costs and drive requests for digital tools that realistically estimate a building's life cycle costs performance. It might be argued that, whilst much work has been done, knowledge-gaps do still remain in research related to accurate cost-prediction, especially to address the full range of construction-industry variables. There remains a need for better cost estimation methodologies and tools; this is a key motivation for this research.

The work presented here builds upon the argument that using simulation and statistical tools can help solve prediction problems and help create construction-cost models (Cheng et al. 2010). For example, a regression-model is one such statistical method towards identifying the impact of factors on construction-costs. Selecting the best regression equation to estimate construction-cost depends on the relations between specific factors and construction-costs.

Simulation and modelling tools such as expert-systems, neural-networks (NNs), fuzzy-logic (FL), and genetic algorithms (GAs) are argued as being able to solve prediction problems (Cheng et al. 2010). Neural-networks (NNs) are considered as potentially important applied-methods to address cost-estimation problems. Artificial-neural-networks (ANNs) are data modelling methods that try to address complicated problems by formulating data relationships; albeit that there is no simplistic equation that can map between data variables.

The estimation methods alluded to above are deemed able to begin to address the issue of absent standardised methodologies for both data collection and analysis in cost-estimation; to extend such work, the research presented here aims to compare the performance of regression, and then neural-networks to estimate the running-cost of building-projects.

In addition, cost-significant items are used to identify the most significant cost-factors, rather than seeking to take into account all cost-factors. This will help model development by easing use and facilitating result(s) comparisons. In construction, typically ~80% of the value (total cost) of a project, might be expected to be contained within ~20% of bill-of-quantities items (Alqahtani and Whyte 2013), typifying Pareto's 80/20 rule. Furthermore, this research takes into account non-cost factors building upon previous studies conducted to identify the main non-cost factors affecting the life-cycle cost-estimation. Finally, cost-significant and non-cost factors are used as input data for both models.

Building upon the rationale above, the structure for the remainder of this discussion involves: firstly, a brief summary of the application of *multiple-regression & artificial-neural-networks* in construction;

then, descriptions of the data sets used in this study; after-which, the approach of both methods is presented and the results for both methods summarised.

Review of past implementation of multiple-regression and neural-networks

This research, seeking improved estimation accuracy, builds upon previous work that confirms that neural-networks and multiple-regression can be employed in economic related areas, providing answers in cases of complicated mathematical calculation and, can be used in data trade analysis and forecasting. Past implementation of both methods towards addressing construction issues such as cost-estimation is discussed briefly below.

In building projects sectors, Boussabaine and Elhag's (1998) work investigated neural network to develop two ANNs models to predict the lowest tender price of primary and secondary school buildings. They concluded that the more significant factors contributed in developing an ANN model, the better the outcomes that will be achieved. Murat and Zeynep (2004) also examined cost estimation for residential buildings by artificial-neural-networking, concluding that a neural-network model can reduce the uncertainties in estimating a structural system of building. Kore et al (2005) conducted a study to compare the performance of the three estimating methods (regression, case based reasoning and ANN) in estimation construction cost and, concluded that neural networks models gave more accurate prediction results than multiple regression or case based reasoning models, albeit that the neural network models are unable to provide an explanation due to inherent 'black box' techniques. Also, they found that regression models have no specific approach that will help estimators select the cost model that best fits input and output variables to a given best accuracy of estimation. Two years later, Wheaton and Simonton (2007) developed hedonic cost models for residential and office properties focusing on 'true' trends and analysing cost/building activity correlation; cost-indicators/ and drivers, whilst were not central to Wheaton, direct similar *semilog* regression-models in this present study presented here.

In Germany, based on 70 German residential properties, relevant building construction cost drivers were identified (Stoy, Pollalis, & Schalcher, 2008). They reviewed previous studies' supplies to list cost drivers and then re-examined these via interviews with expert practitioners. They concluded that compactness of the building, number of elevators, size of the project, expected duration of construction, proportion of openings in external walls, and region are the most important variables affecting building construction cost. This study supports arguments that (re)identification of variables affecting the construction cost estimation process need explicit recognition.

In Egypt, Elkassas et al (2009) also conducted research to predict construction project costs using neural-network modelling, creating 3 back-propagation neural-networks to develop a model. They found that neural network modelling provided a good, accurate estimation.

Chang (2010) also sought to predict maintenance costs developing 4 neural-networks-models, comparing result(s) with regression to identify which model has the least error. Four main factors were argued to affect costs (age, floor, classroom, and elevator number). Four models were classified based on the number and type of variables in each model, concluding that the prediction accuracy of ANNs modelling is better than a regression. Arafa et al. (2011) also developed a neural-network model to predict construction-project cost, concluding that neural-networks can estimate building-cost without detailed design.

In the highway construction sector, Hegazy and Ayed (1998) created a simple ANNs model to develop parametric cost estimating for highway-projects, towards determining network weighting by simplex optimisation and genetic-algorithms (GAs) and, back propagation towards training process

optimisation, concluding that an ANNs improved estimate accuracy, over multiple regression analysis. Similarly, Al-Tabtabai (1999) reviewed 40 highway projects to develop an ANNs model to estimate the percentage increases in the cost from a baseline, with nine variables using input layers. This model generated a mean absolute percentage error of 8.1%. Sodikov (2005) also used neural-network approaches to predict costs in highways arguing that error in multiple regression was higher than ANNs.

In bridging construction sector, Creese and Li (1995) sought prediction of timber bridge costs using an ANNs model, finding that accuracy of estimation of total cost increased as the input variable used increases, concluding that prediction accuracy of neural networks model is superior to regression approaches. This is indicative of the fact that variables affecting the timber bridge costs, other than those used in the research, need to be identified and that, also increasing the input variables lead to decreases in margin error.

For another purpose, Williams' (1994) study into neural-network abilities in estimating a construction-cost index created two back-propagation neural-network models to estimate the change in cost index for one and six month periods, and concluded that the estimated result from neural-networks modelling gave greater error than both regression approaches and exponential smoothing. As far as accuracy is concerned, the research shows that the neural-network modeling is able to resolve disadvantages in the regression approach and provide more accurate results. This study supports arguments that variables affecting a construction cost estimation process must be identified.

Another piece of work conducted by Attalla and Hegazy (2003) to investigate the risky environment of reconstruction projects and identify the significant factors affecting their cost performance, comparing ANN-modelling with regression-modelling concluded that, both produced close predictabilities of cost-deviation. However, the neural-network-modelling technique estimated (similar) results using 18 variables, whilst regression utilised only 5, concluding ANNs as more suitable for high-levels of uncertainty where decision-support is needed. Regarding key variables, Aibinu and Pasco (2008) indicate that, in Australia, the *size* of project is the most influential factor in cost-estimation of construction and that (large) over-estimations are common in small projects and, argue that improving cost-estimation requires probability estimation and simulation of past estimates.

For probability analysis methods there are several methods that can be conducted, with Monte Carlo Simulation perhaps most readily available; albeit that, disadvantages for this method might include: probability subjectively evaluated today may be different in future (Whyte & Scott, 2010); and, that sensitivity analysis does have a goal to quantify risk but rather to identify factors that are risk sensitive and only one parameter can be varied at a time (R. Flanagan & Norman, 1993). Similarly in terms of readily available tools, off-the-shelf computer software applications (stemming from early computer integrated construction (CIC) and BIM initiatives) also continue to split opinion. Researchers (Hu, 2008; Arayici, 2005; Green, 2009 and Waston 2005) have long debated the extent to which LCC has still to be fully integrated into current integrated packages, and that this remains an important part of roadmap towards future development of computer software.

Upon reflection, the studies above find that two techniques - multiple-regression and artificial-neural-networks - can interpret relationships between costs and specific variables. However, disadvantages for both techniques can be flagged explicitly: firstly, in impletion of regression methods, the relationship between variables must be assumed, with numbers of input variables somewhat limited; secondly, literature regards ANNs as a 'black box' approach, where the model is built and utilised without explanation of what the model has learned, thus is suitable if the main objective is only to apply ANNs to make estimations from existing data. To address the first point, this research argues that the main factors affecting cost-estimation accuracy can be identified by using the concept of cost-significant-items CSIs. The second disadvantage (*mentioned by previous research applications and techniques towards opening the 'black box' to explain how ANNs generate estimates*) finds disagreement over the factors affecting final results (Garson 1991; Gevrey et al. 2003; Milne 1995; Olden and Jackson 2002; Recknagel et al. 2006), and perhaps need for Connection Weight (CW)

methods, (Olden and Jackson (2002) which seem to outperform other approaches in assigning relative contributions of input variables in output estimations, goes towards (CW) clarification of black-box output.

Methodology

A comprehensive catalogue of 20 building-projects compiled previously by Al-Hajj (1991) is used to pilot this current study to compare the performance of both estimation methods (multiple regression and ANNs models) to estimate operation and maintenance costs. The data comes from three sources: York University, a facility management company and Building Maintenance Cost Information Services (BMCIS). The main cost factors affecting the accuracy of running cost estimation for these 20 building projects has already been identified by Al-Hajj (1991), employing the concept of cost significant items (CSIs) where 11 items are identified as most important across all buildings studied over the period of 18 years; these items are internal decoration, roof repair, internal cleaning, staff engaged in servicing building (laundry and portorage), management fees, rate, insurance, gas, electricity and fuel oil expenditure. This historical catalogue is give credence by recent readily available reports from the BCIS that, the result of LCC analysis of building projects does show that running costs represent 60% of total LCC and this percentage is increased by approximately 5% during changes in the period of analysis from 18 to 20 years. Similarly updated BCIS data identify operation cost as 70% of the total running cost and that this percentage is decreased by approximately 3% during to changes in the period of analysis from 18 to 20 years. The 11 items identified by Al-Hajj still represent 70% of total running costs during a period of analysis of 18 years. Tax rate represents 24% of total running cost and 35% of total operation costs. Internal cleaning represents 19% of total running cost and 27% of total operation costs. Gas, electricity and fuel oil expenditure represent 10% of total running cost and 14% of total operation costs. Internal decoration is representative of 4% of total running cost and 14% of total maintenance costs. Finally insurance, management fees and staff engaged in servicing building represent between 4%-6% of total running costs.

Indeed the 7 important non-cost factors identified as affecting the building's estimation cost remain: type of building, gross floor area, area of pitched roof, area of flat roof, area of external glazing, number of stories above ground floor and under ground floor.

Input data for both methods consist of eight factors including: 7 non-cost factors and the total value of CSIs. Output data for both methods represents actual values of running-costs over the 20 building projects.

Multiple linear regression-modelling:

Multiple-regression-models are formed as equation (1):

$$\text{Running-costs} = a + b_1 * X_1 + b_2 * X_2 + \dots + b_n * X_n \dots \dots \dots (1)$$

Where: a- intercept b_1 to b_n ; regression-coefficients- X_1 to X_n ; independent variables of all 8 input variables considered independent variables for regression-models.

For evaluation of the multiple regression model, the adjusted R^2 -value and P-value are key; the R^2 -value represents the percentage variability in the costs that can be determined by the variables involved in the model. If R^2 is equal or close to 1, then there is good correlation (good-fit) between the actual value and the estimation model output. Furthermore, In order to improve the result of multiple-regression for fair comparison with ANNs, a significance level (p-value) was used to identify the variables to be eliminated. In general, the variables with p-value close-to or less-than 0.05 are considered to have an important contribution to the model (Ontepeli 2005). The stepwise method will be applied to identify the most important variables. The procedure of this method starts by including all 8 variables (initial experimentation) in the model; then an identification of the p-value for each

variable, if a variable has p-value more than 0.05; and, subsequent elimination until identification of the best model with a consistent variable, p-value equal or less-than 0.05.

Validation of multiple linear regression-modelling

The R²-value (good-fit) of a model is *not always* enough for an accurate estimation; performance of the models should also be assessed by applying cross validation techniques (Sonmez 2008). For this reason, the 20 projects are divided into two sets. One set consists of 17 projects used to ‘build’ the model, with 3 projects used towards model validation. The performance of the model will be evaluated based on the value of mean percentage error (MPE) (equation2) and R² value, such that

$$MPE = \frac{1}{n} \sum_{i=1}^n \frac{O_i - P_i}{O_i} * 100 \dots\dots\dots (2)$$

where: O_i: the actual output; P_i: The model output; i is the project number and n is the total number of projects.

Neural-network modelling

This paper uses spread-sheet optimization (Microsoft Excel’s inherent solver function) to build the ANNs model. This paper applied 7 steps (adapted from suggestions by Hegazy and Ayed (1998)) to build an ANNs model by spread-sheet.

In this paper, the aim is optimization and to decrease the NN weighted error to reach the 0 value. In order to achieve this objective, the adjusted variable has been selected as the weight(s) from inputs to hidden nodes and from hidden nodes to outputs. Optimization constraints were set to limit the proportion error on both training and test project to 3% and 1% or lower to avoid erroneous network result(s) on individual training projects.

Traditional parametric (trial and error) has been applied to select the number of hidden layers and the number of hidden nodes. During the training process, hidden layer numbers and hidden nodes are adjusted to find the best artificial-neural-network model to give the minimum value for the Root Mean Square (RMS) and MPE, equation (2 and 3) for output parameters. Tangent Sigmoid is used as a transfer function of NNs model.

$$RMS = \sqrt{\sum_{i=1}^n \frac{1}{n} (O_i - P_i)^2} \dots\dots\dots (3)$$

RMS: a root-mean-square error, n: number of sample using in the training stage, O_i: the actual output. P_i: The model output.

As with regression modelling above, the 20 projects are divided into two sets: 17 projects used for training the model; and the remaining 3 projects used to test the procedure. As mentioned, one of the objectives of training the model is to identify the best structure of a ANNs model. The acceptable level of the result of the model is evaluated based on the value of RMS (equation3) and MPE (in equation 2).

Connection weight method (CW)

The connection weight(ed) method was applied in order to rank the importance of input (eight) variables in predicting the output variables (running costs) for neural-networks. This method (Olden et al. 2004)) calculates the sum of products of weights of the connection from input nodes to the hidden nodes and the weight(ing) of the connection from hidden nodes to the output node for all input variables. The larger the sum of the connection weight(ing), the more important is the corresponding input variable. The relative importance of input variable i can be defined as:

$$RI_i = \sum_{N=1}^n (W_{iN} * WN_o) \dots\dots\dots (4)$$

Where: RI_i : the relative importance of input variable I; n is the total number of hidden neuron; W_{iN} : the weight of connection from input neuron I and hidden neuron N; and, W_{NO} : the weight of the connection between hidden neuron N and output neuron.

Paired t-Test:

In order to create statistical significance, a statistic test is suitable to comparatively analyse estimation accuracy between ANNs and multiple regression for predicting running costs. As the data used in this study in both models are same, paired t-test (two samples for mean) was applied to estimation accuracy (absolute mean percentage error) to test all hypotheses (paired t-test by SPSS software), where hypotheses are proposed such that:

- a) H_0 : there is no prediction accuracy difference in regression and NNs methods ($\mu_1 = 0$, Absolute mean percentage error is close to zero)
- b) H_1 : There is prediction accuracy difference in regression and NNs methods ($\mu_1 \neq 0$, Absolute mean percentage error is to large)

Results

Artificial-neural-network models

The 6 model trials and error were applied to find the number of hidden nodes on hidden layers. It was obvious that increasing the number of nodes in hidden layers leads to changing the value of both RMS and Absolute Difference error. 4 hidden nodes provided the lowest RMS value of 0.011 and an absolute difference error value of 0.29%.

The neural-network model developed in this paper consists of 3 layers: input layer of 8 nodes, one hidden with 4 nodes (trial and error in a training stage was applied to determine number of nodes) and output-layer contained in one node (the total running cost).

It is shown from the results that the running-cost model developed by ANNs modelling performs well; no important differences are recognised between the estimated and actual running-costs. Results are tabulated below in Table 1. The MPE result from neural-network modelling for the testing projects was 0.179 %. The expected accuracy of the ANNs model at training and testing stage is described in table 1 below.

The ANNs model is able to estimate the total running cost with an average accuracy of 99%. The neural-network model results from both training and testing stages and the actual value of running costs were passed to regression analysis in order to investigate the model response in more detail. The result of linearly regression is presented graphically in figure 1 below.

In training and testing stages, the R^2 is close to 1, indicating a good fit and linear correlation between the actual running-cost and the ANNs result at training and testing stages.

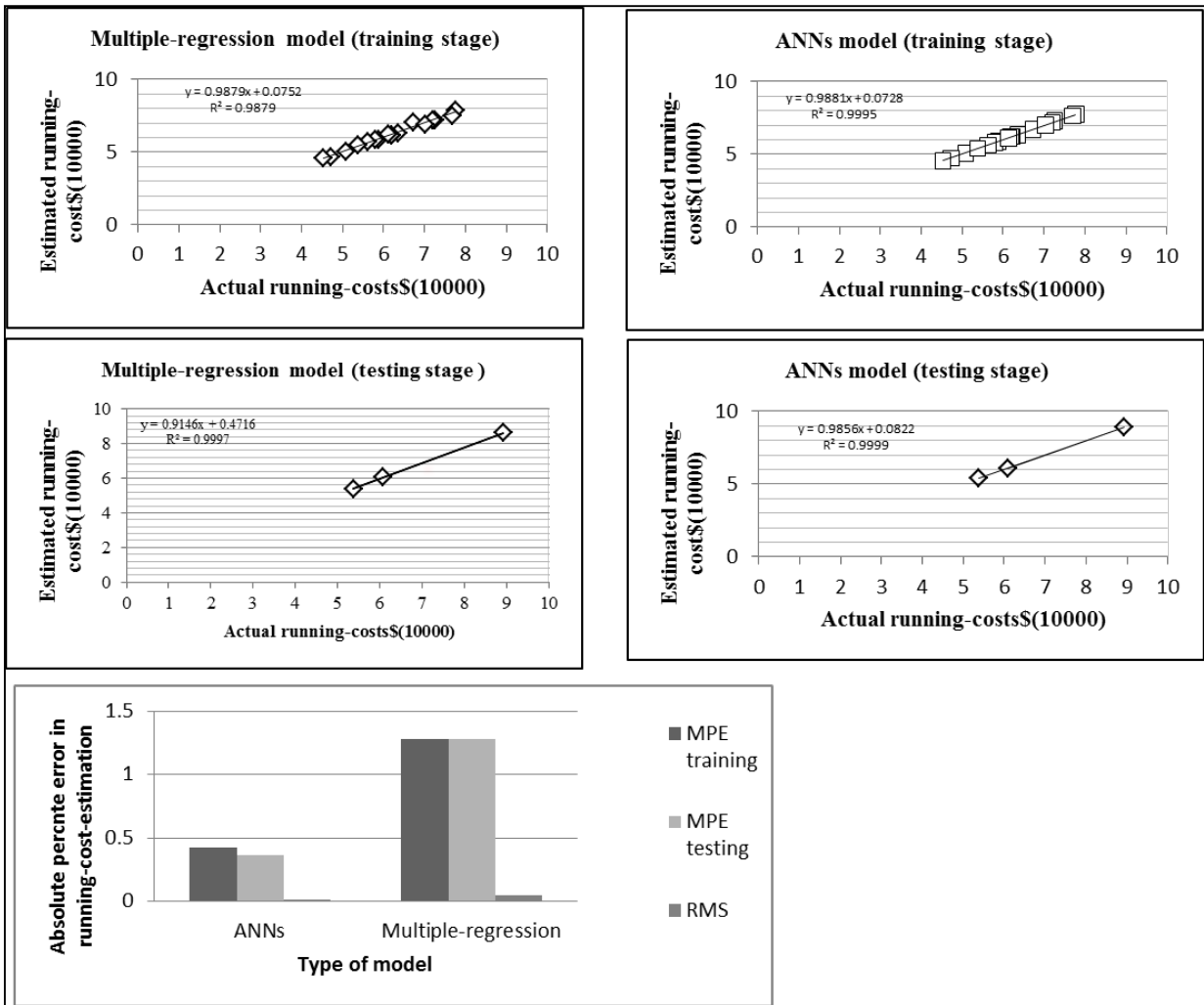


Figure 1 Comparisons the result of adjusted R^2 , MPE and RMS for two models at each stage.

Projects	Stage	Actual running cost	ANNs		Multiple-Regression	
			Prediction	different error [(Actual-prediction)/Actual]*100	Prediction	different error
1	Training set	63550.86	63495.95	0.086	62830.42	1.133
2		58834.79	58833.92	0.001	58834.79	0
3		72589.29	72952.23	-0.500	72167.02	0.581
4		58006.39	58009.13	-0.005	58303.04	-0.511
5		77712.42	77717.36	-0.006	78385.60	-0.866
6		67320.86	67013.82	0.456	70108.44	-4.140
7		62094.77	62103.54	-0.014	61437.80	1.058
8		61905.1	62199.50	-0.476	61474.89	0.694
9		72134.2	71773.57	0.500	72009.06	0.173
10		56257.54	56258.64	-0.002	56853.24	1.058
11		50832.14	50830.18	0.004	50415.96	0.818
12		61296.04	60989.67	0.500	62304.16	-1.644
13		76918.18	76533.59	0.500	75166.71	2.277
14		70330.57	70108.58	0.316	69014.98	1.870
15		53760.86	54029.55	-0.500	55004.36	-2.313
16		47220.45	47456.55	-0.500	46234.51	2.087
17		45350.81	45577.54	-0.500	45570.20	-0.483
18	Test set	60832.72	60832.69	0.000	60673.16	0.262
19		89369.22	88890.31	0.536	86394.96	3.328
20		53834.61	53834.61	0.000	53701.51	0.247
Mean percentage error (testing)			0.179		1.279	
Expected Accuracy						
Models	Stage	1- Mean percentage error (MPE)	2- Standard Deviation(ST)	3- Accuracy range (MPE±ST)		
ANNs	Training stage	-0.008	±0.378	(-0.386 to +0.370)		
	Testing stage	0.179	±0.309	(-0.131 to +0.488)		
Multiple-regression	Training stage	-0.019	±1.668	(-1.687 to +1.649)		
	Testing stage	1.279	±1.774	(-0.495 to +3.053)		

Table 1 the main results for both models

Relative importance of independent variables in case of ANNs:

As mentioned one disadvantage of ANNs modelling is the (black box) lack of illustration for the relative importance affecting the independent variables. However, CW is applied to discover the relative importance affecting independent variable(s) (input data) to the dependent variables (output data). It is clear that ‘CSIs’ are most important variable(s) influencing running cost value(s); number of storeys underground floor is the second most important variable, followed by project type; area of external glassing; number of storeys above ground floor; area of pitched roof and gross floor area with area of flat roof of lesser importance.

Multiple-regression-model

In this case study, Multiple-regression-modelling was developed to compare results with neural-network modelling. Multiple-regression-models are formed as equation (1).

Similar to ANNs model, 17 projects were used to develop the regression-model. Initial experimentation with multiple regression models include the 8 variables used in developing the ANNs models to clarify that there are insignificant variables which need elimination to improve the result of the model. For example in initial experimentation 'number of stories above ground floor' in an insignificant factor, based on p-value (0.668).

In order to improve the result of multiple-regression-modelling for fair comparison with ANNs modelling, significant level (p-value) was used to identify the variables to be eliminated. In general, the variables with p-value close to or less than 0.05 are considered to have an important contribution to the model (Ontepeli 2005). This method is termed a stepwise method. The procedure of this method starts by including all 8 variables (initial experimentation) in the model; then identifies the p-value for each variable; if a variable has p-value more than 0.05 then it will be eliminated until identification of the best model with consist variable(s) with p-vale equal or less than 0.05.

The influence of each independent variable (of input data) in regression modelling can be identified by checking the value of coefficient and p-value for each variable. The initial model considers two variables as the most important factor. They are, in order of importance: CSIs and project type. The relationship between running costs and these variables are positive. After an applied stepwise method and identification of the best model, six variables are highlighted as most influential, in order of importance: CSIs, area of flat roof, project type, and gross floor area, number of storeys underground floor, and area of pitched roof. Only the area of flat and pitched roof has a negative relationship with running-costs. It is clear that the stepwise method results lead to a very different conclusion(s) from the initial model.

The best regression-model can thus far be represented with the following equation:

$$\text{Running-costs} = -0.155 + 0.253 * \text{gross-floor-area} - 0.097 * \text{area of pitched roof} - 0.382 * \text{area of flat roof} + 0.095 * \text{number of storeys under ground floor} + 0.891 * \text{CSI} + 0.194 * \text{project type}$$

From the table (1), the MPE result from multiple regression modelling for the testing projects was 1.279. The expected accuracy of the neural-network model at training and testing stage is introduced also in table (1).

The multiple-regression-model is thus far, able to estimate the total-running-cost with an average accuracy of 97%.

Paired t-Test: regression Vs NNs modelling

The result of paired t-test indicate that the P-vale <0.05, H₀ is rejected; this means the Absolute mean percentage error of regression modelling was significantly more than the neural-network model (approximately three times), indicating no similarity between the results of both models. Therefore it is argued that ANNs modelling gives more effective results than regression modelling alone.

Discussion:

Currently, building cost-estimation is complex and time consuming. Based on the results generated by this study's models, cost-estimation function(s) may now go towards an appropriate/suitable cost estimate in a much reduced time-frame (than existing approaches).

From the result presented above, it is clear that ANNs models provide excellent performance in terms of both training and testing projects when compared to multiple-regression-modelling; figure (1) illustrates the MPE at training and testing stage for both models with RMS. The best overall model developed here by ANNs is able to estimate running-costs with 99% accuracy.

These results also show that an ANNs may be an attractive substitute for regression because the estimation process becomes more complex as the dimensionality of the cost drivers set grows and the neural networks are capable of accepting a larger number of potential cost drivers than regression. In addition the method of selecting a best model for neural network(s), dependent upon a learning curve, carries more worth than statistical models where decisions are fewer and guidance is readily available from texts and software. Furthermore, the relationship between the dependent variables and independent variables in regression modeling must be reviewed in advance.

Both methods presented above are able to identify 'importance factors' affecting an estimation of running costs. It is interesting to note that both methods provide the same result regarding the most important variable influencing estimation results, namely CSIs.

Results here show that building type has an effect on LCC, with total cost of design, construction, maintenance and operation dependent upon the nature of asset specialisation and relative material, validation gained somewhat by literature (Al-Hajj (1991) where models of specific building types have a higher predictive accuracy. Roof type too influences cost estimation model outcomes, with flat roofs (empirically found to be) more expensive than pitched-roofs of comparable quality due to simplicity in spanning large areas with roof trusses rather than deep beams. Similarly gross floor area is important with large building areas requiring comparable larger resources beyond the design, construction phases to operate and maintain. Important factors identified by this research, reconfirm previous work studies highlighting project nature, roof type and gross floor area as key to an estimation of cost (Aibinu and Pasco 2008, Cheng et al. 2010, Elhag and Boussabaine 1998, Sonmez 2008, , E.M. Elkassas et al. 2009). Further research is also appropriate to identify (through literature, historical data) supplementary (significant) cost driver(s) affecting cost estimation through detailed analyses of project location, inflation rate, and project design flexibility towards potential life cycle increases.

Conclusion:

Cost estimation has a very significant role in project decision-making and can contribute to enhanced design parameters by improving decisions related to best fit specifications, and effect customers' satisfaction levels by advising of potential reductions in projects cost, as well as help timely project delivery of fit for purpose components onsite. Multiple-regression (MR) and ANNs modelling were used here in order to improve the quality of the estimation process. Identification of the main factors affecting the accuracy of an estimation of (whole) cost is an important first step in the research presented. Resultantly, using the concept of cost significant items (CSI) alongside an application of historical project cost data, the model developed and created by this work (using input data from 20 previously catalogued building-projects) allowed a better estimate of the total running cost of construction assets. Regression models have no clear defined method to help estimators select the best design model that can be used to find the relationship between independent and dependent variables. ANNs *do* seem to be very flexible and *can* be adopted easily across different designs. ANNs *are* able to accept a larger number of independent variables than regression.

This research developed then compared the performance of estimating approaches MR and ANN. The results of this comparison reveal that ANNs modelling is the best (most accurate means and) method to estimate running-costs. The model developed in this paper, for running costs estimation of construction projects, is presented as a way to help enhance accuracy levels of future project cost estimation.

References:

- Aibinu, A. A. and Pasco, T. (2008) "The accuracy of pre-tender building cost estimates in Australia", *Construction Management & Economics*, Vol. 26 No. 12, pp. 1257-1269.
- Akintoye, A. and Fitzgerald, E. (2000) "A survey of current cost estimating practices in the UK", *Construction Management and Economics*, Vol. 18 No. 2, pp. 161-172.
- Al-Hajj, A. N. (1991) 'Simple cost-significant models for total life-cycle costing in buildings, unpublished thesis' Ph.D. thesis, University of Dundee (United Kingdom).
- Alcabes, J. (1988) "Organizational concept for a coordinated estimating, cost control, and scheduling division", *AACE*.
- Alqahtani, A. and Whyte, A. (2013) "Artificial neural networks incorporating cost significant Items towards enhancing estimation for (life-cycle) costing of construction projects", *Australasian Journal of Construction Economics and Building*, Vol. 13 No. 3, pp. 51-64.
- Attalla, M. and Hegazy, T. (2003) "Predicting Cost Deviation in Reconstruction Projects: Artificial Neural Networks versus Regression", *Journal of Construction Engineering and Management*, Vol. 129 No. 4, pp. 405-411.
- Carr, R. (1989) "Cost-Estimating Principles", *Journal of Construction Engineering and Management*, Vol. 115 No. 4, pp. 545-551.
- Chang Sian, L., Pei Jia, C. and Sy Jye, G. (2010) "Application of back-propagation artificial neural network to predict maintenance costs and budget for university buildings", *IEEE*, pp. 1546-1551.
- Cheng, M.-Y., Tsai, H.-C. and Sudjono, E. (2010) "Conceptual cost estimates using evolutionary fuzzy hybrid neural network for projects in construction industry", *Expert Systems with Applications*, Vol. 37 No. 6, pp. 4224-4231.
- Creese, R. C. and Li, L. (1995) "Cost estimation of timber bridges using neural networks", *Cost Engineering*, Vol. 37 No.5, pp. 17-17.
- Dysert, L. and Elliott, B. G. (2002) "The Estimate Review and Validation Process", *Cost Engineering*, Vol. 44 No. 1, pp. 1-17.
- E.M. Elkassas, H.H. Mohamed and H.H. Massoud (2009) "The neural network model for predicting the financing cost for construction projects", *International Journal of Project Organisation and Management*, Vol. 1 No. 3.
- Elhag, T. M. S. and Boussabaine, A. H. (1998) "An artificial neural system for cost estimation of construction projects", *Association of Researchers in Construction Management*, pp. 219-226.
- Garson, G. D. (1991) "Interpreting neural-network connection weights", *AI Expert*, Vol. 6 No. 4, pp. 46-51.
- Gevrey, M., Dimopoulos, I. and Lek, S. (2003) "Review and comparison of methods to study the contribution of variables in artificial neural network models", *Ecological Modelling*, Vol. 160 No. 3, pp. 249-264.
- Hashem Al-Tabtabai, A. P. A. M. T. (1999) "Preliminary cost estimation of highway construction using neural networks", *Cost Engineering*, Vol. 41 No 3, pp. 19-24.
- Hegazy, T. and Ayed, A. (1998) "Neural Network Model for Parametric Cost Estimation of Highway Projects", *Journal of Construction Engineering and Management*, Vol. 124 No. 3, pp. 210-218.

- Kim, G.-H., An, S.-H. and Kang, K.-I. (2004) "Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning", *Building and Environment*, Vol. 39 No. 10, pp. 1235-1242.
- Milne, L. K. (1995) "Feature selection with neural networks with contribution measures", in *the Australian Conference on Artificial Intelligence AI'95*, Canberra,
- Murat Günaydın, H. and Zeynep Doğan, S. (2004) "A neural network approach for early cost estimation of structural systems of buildings", *International Journal of Project Management*, Vol. 22 No. 7, pp. 595-602.
- Olden, J. D. and Jackson, D. A. (2002) 'Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks', *Ecological Modelling*, 154(1-2), 135-150.
- Olden, J. D., Joy, M. K. and Death, R. G. (2004) "An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data", *Ecological Modelling*, Vol. 178 No. 3-4, pp. 389-397.
- Ontepeli, M. (2005) "Conceptual cost estimating of urban railway system projects", unpublished thesis Middle East Technical University (METU).
- Recknagel, F., Cao, H., Kim, B., Takamura, N. and Welk, A. (2006) "Unravelling and forecasting algal population dynamics in two lakes different in morphometry and eutrophication by neural and evolutionary computation", *Ecological Informatics*, Vol. 1 No. 2, pp. 133-151.
- Sodikov, J. (2005) "cost estimation of highway projects in developing countries: artificial neural network approach", *The eastern Asia Society for Transportation Studies*, Vol. 6, pp. 1036-1047.
- Sonmez, R. (2008) "Parametric Range Estimating of Building Costs Using Regression Models and Bootstrap", *Journal of Construction Engineering and Management*, Vol. 134 No. 12, pp. 1011-1016.
- Stoy, C., Pollalis, S. and Schalcher, H. (2008) "Drivers for Cost Estimating in Early Design: Case Study of Residential Construction", *Journal of Construction Engineering and Management*, Vol. 134 No. 1, pp. 32-39.
- Wheaton, W. C. and Simonton, W. E. (2007) "The secular and cyclic behavior of 'true' construction costs", *The Journal of Real Estate Research*, Vol. 29 No. 1, pp. 1-25.
- Williams, T. P. (1994) "Predicting Changes in Construction Cost Indexes Using Neural Networks", *Journal of Construction Engineering and Management*, Vol. 120 No. 2, pp. 306-320.
- Yaman, H. and Tas, E. (2007) "A building cost estimation model based on functional elements", *A, Z ITU Journal of Faculty of Architecture*, Vol. 4 No. 1, pp. 73-87.