# An Artificial Neural Network Approach to Structural Cost Estimation of Building Projects in the Philippines

Roxas, Cheryl Lyne C.[1], Ongpeng, Jason Maximino C.[2]
[1,2] *Civil Engineering Department, De La Salle University Manila, 2401 Taft Avenue, Manila, Philippines,*
**Corresponding Author: cheryl.capiz@dlsu.edu.ph*

**Abstract:** The success of any project undertaking is defined by improved quantity and cost estimation techniques that will facilitate effective cost and time control in projects. The objective of this study is to develop an artificial neural network (ANN) model that can predict the total structural cost of building projects in the Philippines. Data from thirty building projects were collected and randomly divided into three sets: 60% for training, 20% for validating the performance and 20% as a completely independent test of network generalization. Six input parameters, namely: number of storeys, number of basements, floor area, volume of concrete, area of formworks, and weight of reinforcing steel. These variables were entered into the ANN architecture and simulated in MATLAB. The feedforward backpropagation technique was used to generate the best model for the total structural cost. The best ANN architecture consists of six input variables, seven nodes in the hidden layer and one output node. The resulting ANN model also reasonably predicted the total structural cost of building projects with favourable training and testing phase outcomes.

**Key Words:** structural cost; artificial neural network; ANN

## 1. INTRODUCTION

Cost is probably the first to be considered when it comes to construction projects. A good estimate of quantities and costs in a construction project is a vital factor in its success. Because of the complexity of the construction industry and uniqueness of every project undertaking, several factors may affect the overall project cost.

A number of items, such as the structural, architectural, sanitary, electrical and air-conditioning system works determine the total cost of buildings. Olotuah (2002) observed that 60% of the cost of residential buildings can be attributed to the cost of building materials. Meanwhile, the structural frame covers 25% of total construction cost in multi-storey reinforced concrete residential buildings (Gould and Joyce, 2000).

The structural system works form the bulk of the overall cost of buildings. Therefore, utmost care must be exercised in the design of structural systems if a considerable reduction in cost is desired. However, questions concerning the reliability and accuracy of available cost estimating techniques are present. Therefore, there is a need to address this issue by introducing a new and alternative approach to cost estimating.

Oftentimes, the ordinary least squares regression approach is applied and the model is selected based on the coefficient of determination, $R^2$. However, because of high correlation among a large group of variables, this method tends to generate regression coefficients estimators that will badly perform in the presence of multi-collinearity (Tewari and Singh, 1996). Moreover, the variance of the ordinary least squares estimator become inflated, which results in the low possibility of the estimator being close to the true value of the regression coefficient (Montgomery and Peck, 1982). This can be improved by determining uncorrelated variables to be included in the regression model.

Many quantity and cost estimation models have been developed. Multiple linear regression is a very useful statistical tool in analyzing and predicting the contribution of a potential new item to the overall estimate. Other methods have been introduced and applied for cost estimation such as principle component analysis, case-based reasoning and ANN.

The purpose of this work is to explore the ANN technique and determine the total structural cost of buildings. To accomplish this goal, structural cost data were collected and were used simulated in MATLAB.

This work is organized as follows. First, a review of different cost estimating techniques and ANN applications was conducted. Next, the data collection and preparation and ANN modeling with the input and output variables, architecture, trials and error analysis are presented. Next, the final model with its architecture follows. Lastly, results are summarized and future studies are recommended.

## 2. LITERATURE REVIEW

The goal of construction management is to efficiently and economically apply the required resources to construct a facility of acceptable quality within the time specified, approved budget and without compromising safety. Among these major criteria, cost is often considered first by most project managers. Cost management is therefore critical in any project undertaking. It is a process involving cost planning, estimating, budgeting and controlling so that project complies with the approved budget (Han, 2008). Cost forecasting depends on limited, noisy and approximate information during the pretender stage of a project which make it difficult to understand its underlying cost drivers (Aibinu et.al., 2011).

There are several cost estimating methods, namely: traditional detailed breakdown cost estimation, average estimation per construction area, comparative cost estimation, multiple linear regression, principal component regression, and case-based reasoning (CBR) (Manprasert, 1998; Arafa and Alqedra ,2011; Gunaydin and Dogan, 2004; Chan and Park, 2005; Kim G.H. et.al, 2004; Han 2008; Kim D.Y., 2008).

Traditional cost estimation techniques become inaccurate ad impossible to implement when the data are ambiguous and subject to change (Arafa and Alqedra, 2011). Linear relationship between the final cost and basic design variables is assumed in comparative cost estimation (Gunaydin and Dogan, 2004). According to Arafa and Alqedra (2011), the linear relationship in regression limits the capability to fit data and may not always be appropriate for it entails a functional relationship. The relationship between the input and output variables in regression models are straightforward and too simple compared to the complexity of the real world relationship between those variables (Aibinu et.al, 2011). Disadvantages of regression estimating models for construction cost include: no specific of clearly defined approach that will help estimators choose the cost model that best fits historical data to a given cost estimating application (Garza and Rouhana, (1995); Adeli and Wu, (1998); Bode, (1998); Bode, (2000)); a certain type of multiple equations and its data are assumed to be suitable for the regression equations (Adeli and Wu, (1998); Bode, (1998); Bode, (2000); and variables influencing the estimation must be reviewed in advance and it is also difficult to use a large number of input variables (Bode, (1998); Bode, (2000); Smith and Mason (1997)).

CBR is a problem solving paradigm that is able to utilize the specific knowledge of previously experienced cases and reuse it in a new problem situation (Aamodt, 1994). CBR can generate good quality solutions but may not be effective in situations on naïve adaptation models (Cunningham, 1998). Principal component analysis is more applicable for reducing large number of variables in estimation (Ganiyu and Zubairu, 2010).

With the development of software technology, several approaches to cost estimation techniques have been introduced. One of these is the Artificial Neural Network (ANN). ANN method is the more appropriate method to use for the following reasons: (a) it is able to determine the interdependencies between data when considering significant variables in construction, (b) can deal with non-linear relationships, (c) can handle incomplete data sets and (d) does not depend on assumptions about functional form, probability distribution and smoothness ((Arafa and Alqedra ,(2011), Emsley, et.al., (2002), Camargo et.al., (2003)). Kim G.H. et.al, (2004) investigated different cost estimation models and found that neural networks generated more accurate estimating results than CBR and multiple regression analysis. Since ANN can deal with non-linear relationships, good cost estimating relationships between cost and other significant variables can be eliminated (Kim G.H. et.al, 2004). Moreover, ANN can analyze relationships between cost and cost influencing variables even if the nature of these relationships is unknown because of its ability to learn from examples and capture subtle functional relationship among the data (Aibinu et.al, 2011).

The brain-like structure of neural network enables the models generated to activate a function based on "only the strong survive (Turochy, et.al., 2001). Turochy et.al. (2001) added that the repetitive use and education of the neural network can be easily utilized once a model is established because of its fill-in-the-blank capability. ANN models incorporate contingency reserve to the estimate to reflect unforeseen costs between the early estimate and projected completion time of the project (Aibinu et.al, 2011).

The cost of a construction project is influenced by several factors such as the structural, architectural and engineering systems. Structural cost depends on the area, number of storeys and basement, concrete, formwork, reinforcing steel, post-tensioned area, piles, etc. The architectural cost, on the other hand, depends on the design and quality of materials used for the floor, wall, ceiling, doors and windows, painting, etc. Sanitary, electrical, air-conditioning and elevator systems make up the total engineering system cost.

The value of each item depends on the type and complexity of the design of structure. Mohamed ad Celik (2002) indicated that the design of a building and material selection significantly affects its cost. Fang and Froese (1999) established the relationship of the cost of concrete and formwork to the cost estimation of high performance concrete in high-rise commercial buildings. Design variables like area, perimeter, height and load were used by Squeira (1999) to estimate the cost of low-rise structural steel buildings. A study by Arafa and Alqedra (2011) revealed seven (7) key parameters influencing the structural cost of building construction projects, namely: ground floor area, typical floor area, number of storeys, number of columns, type of footing, number of elevators and number of rooms. The most effective parameters in Arafa and Alqedra (2011) based on sensitivity analysis are the ground floor area, number of storeys, type of foundation and number of elevators in the building. Using principal component regression, Ganiyu and Zubairu, (2010) presented the factors affecting the project cost. Among these are factors related to the adequacy of contractor's plant and equipment, contractor's experience on similar type of project, time allowed for project bid to be evaluated, level of technological advancement and client commitment to timely completion of the project, percentage of repetitive work, level of design complexity, importance for project to be delivered, project scope, percentage of special issues, communication among project team, level of construction complexity, contractor experience on similar size of project and contractors prior working relationship with clients.

Neural networks have been found to be successful in generating reliable and accurate results even with imprecise data. Gunaydin and Dogan (2004) argued that neural networks are capable of reducing uncertainties in the estimate of the structural system of building.

Continuous studies on how to accurately predict the cost of construction projects using ANN is the motivation of the current study. The objective of this paper is to construct, train, validate and test an ANN model that can estimate the total structural cost of building projects.

## 2. METHODOLOGY

### *3.1. Data Collection and Identification of Input Variables*

Structural data from a total of thirty (30) building projects were collected from different construction companies in the Metro Manila. The structural costs obtained were for bidding purposes. Six variables namely: *number of storeys, number of basements, floor area, volume of concrete, area of formworks, and weight of reinforcing steel* were considered. The number of storeys ranges from 5 – 37 floors while the number of basements is between 0 – 3. The total structural cost of buildings ranges between PHP 2-448 million. There were missing values in the data provided but as mentioned in

literature, ANN is capable of analysing incomplete data sets. Also, the period of data collection as well as the time at which the cost estimation of the buildings was done, was not taken into consideration in this study. Doing so could have improved the results as inflation rates affect the cost of materials.
Arafa and Alqedra (2011) considered seven (7) factors affecting the structural system cost, namely: ground floor area, typical floor area, number of storeys, number of columns, type of footings, number of elevators and number of rooms. The total area will reveal the type and thickness of slabs, beam sizes and their costs while the number of storeys will determine the vertical section area of the structural frame and the cost of beams (Arafa and Alqedra 2011)

## 3.2. Building the ANN Models

The *Neural Network Toolbox* in MATLAB R2010a was used in constructing the ANN total cost model estimation. The total cost and corresponding parameter data from 30 building projects were divided into three sets: 60% for training the neural network, 20% for validation and 20% for testing. These sets were randomly selected in MATLAB.

In this paper, the tan-sigmoid transformation function was applied in the hidden layer and a linear transformation function was used in the output layer, similar to Arafa and Alqedra (2011).

The tan-sigmoid function returns values in the range of -1 and +1. Before training, the data were first normalized to this range to speed up the training process and to improve the network performance. Equation 1 presents the formula used to normalize the training set so that they fall within the range of [-1, +1].

$$y = \frac{y_{\max} - y_{\min}(x - x_{\min})}{(x_{\max} - x_{\min}) + y_{\min}} \qquad \text{Eq. 1}$$

*Where: $y$ = normalized value; $y_{max}$ = maximum value (+1); $y_{min}$ = minimum value (-1); $x$ = original value; $x_{max}$ = maximum value of the data under normalization; $x_{min}$ = minimum value of the data under normalization*

The feedforward backpropagation technique was used to generate the best model for the total structural cost. The backpropagation algorithm gradually reduces the error between the model output and the target output by minimizing the mean square error (MSE) over a set of training set (Gunaydin and Dogan 2004). The MSE is a good overall measure of the success of the training process (Al-Tabtabai, H., et.al., (1999); Turochy, et.al., (2001)) The MSE is expressed as (Eq. 2):

$$MSE = \frac{\sqrt{\sum_{i=1}^{n}(x_i - E(i)^2}}{n} \qquad \text{Eq. 2}$$

*Where: $n$ = number of samples in the training set; $x_i$ = model output related to the sample $i$; $E(i)$ = target output, i.e. total structural cost*

The weights and bias values are updated according to the Levenberg-Marquardt network training function. This is often the fastest backpropagation algorithm and highly recommended, though it requires more memory that other algorithms (Neural Network Toolbox).

## 3.3. Trial Models

The best ANN model to estimate the total structural cost of buildings was determined by defining the number of neurons (nodes) in the input and output layers, number of hidden layers and the number of neurons in each hidden layer.

The model generated utilizes the six input variables. There is no specific rule in determining the number of hidden layers and the number of neurons in each hidden layer (Shtub, A. and Versanob, R., 1999). For simplicity of the current problem, one hidden layer was used and the following rules were employed to determine the optimum number of neurons for a network with 14 and 3 input variables (Oreta 2012):

i. A network with n-input and m-output units requires a hidden layer with at most 2n+1 units (Hecht-Nielson 1998) → **(13 neurons)**

ii. Should be between the average and the sum of nodes on the input and output layers → neurons **(4-7 neurons)**

iii. Seventy-five percent (75%) of the input nodes → **(5 neurons)**

## 3. RESULTS AND DISCUSSION

After several trials, **ANN Structure 6-7-1** (6-input variables, 7-nodes in the hidden layer, 1-output) was found to be the best model to estimate the total structural cost of building projects in the Philippines. Figure 1 shows the ANN Structure 6-7-1. ANN Structure 6-7-1 obtained a satisfactorily acceptable correlation coefficient, R, values of 0.96812, 0.70199, and 0.9548 for training, validation and testing phases, respectively. Figure 2 presents the regression line for ANN Structure 6-7-1. The model developed satisfied the range of nodes in the hidden layer mentioned above. The resulting MSE is equal to $2.98 \times 10^{15}$, which is the smallest value among all model tested (between 4-13 nodes as suggested by Oreta (2012).

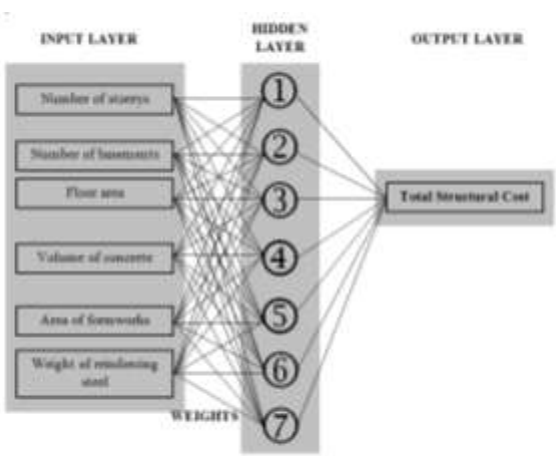


Fig. 1. ANN Structure 6-7-1

An agreement between the actual and predicted values can be seen, especially in the training and overall regression lines, as indicated by the concentration of predicted values around the 45° degree line.

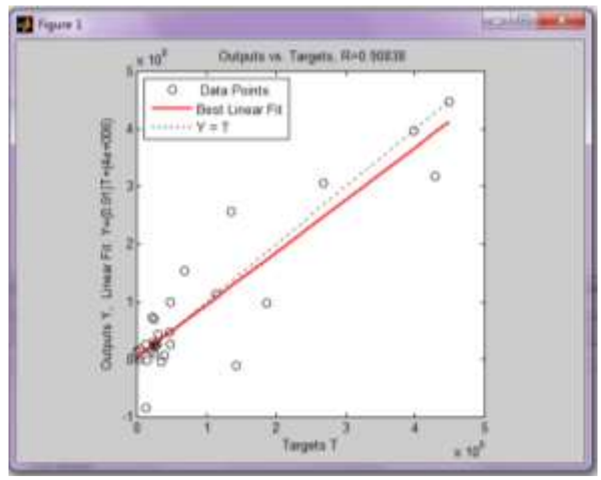


Fig. 2. Regression line for ANN Structure 6-7-1

Figure 3 shows the performance curve of the developed model. The best validation performance is 0.15939 and occurred at epoch 1000. The test set error and the validation set error have relatively similar characteristics.

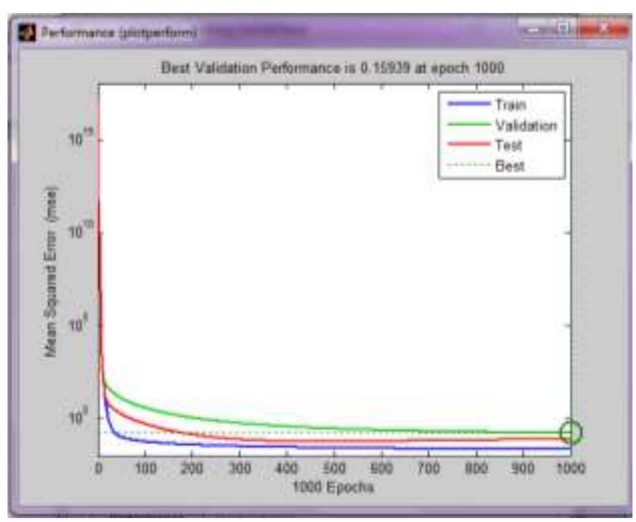


Fig. 3. Performance curve of ANN Structure 6-7-1

This study is different from previous ones in terms of the input parameters used and different construction industry setting. Differences in cultures may affect the methodology employed in building construction, which in turn affects the cost of materials. However, the model generated in this study showed a lower R and higher MSE compared to previous research. The model can be further improved by obtaining a larger data set and improving the MATLAB program for generating the best architectural model.

## 4. CONCLUSIONS

Questions concerning the reliability and accuracy of available cost estimating techniques need to be addressed by introducing a new and alternative approach. The main objective of this work was to apply the artificial neural network technique in estimating the total structural cost of building projects in the Philippines. Data from a total of 30 building projects were collected from different construction companies. Six input variables namely: number of storeys, number of basements, floor area, volume of concrete, area of formworks, and weight of reinforcing steel were considered and further analysed.

ANN Structure 6-7-1 was chosen as the best architecture with the highest correlation coefficient and lowest MSE.

Finally, it can be concluded that the neural network approach provided good predicting ability despite the non-uniform distribution and incompleteness of the data set. Expanded data set is recommended to improve the results. In addition, further statistical analysis can be performed like considering the type of building, i.e., whether low rise, mid-rise and high-rise. Extensive manipulation of the data to be used the MATLAB program can also be explored to improve the results. Sensitivity analysis and relative importance of the input variables can be conducted to enhance the reliability and validity of the results. The models presented can also be compared with other models generated by different ANN softwares.

## 5. ACKNOWLEDGMENTS

The authors would like to acknowledge Mr. Jordan Ng and Mr. Christian Tan, undergraduate civil engineering students, for the assistance in data collection.

## 6. REFERENCES

Aamodt, A., (1994) "Case-based reasoning: Foundational issues, methodological variations and system approaches", *AI Communications*, Volume (7) Nr.1: 33-59.

Adeli, H. and Wu, M. (1998) "Regularization Neural Network for Construction Cost Estimation", *Journal of Engineering and Management*, Volume (124)1: 18-24.

Aibunu, A. et.al., (2011) "Use of artificial intelligence to predict the accuracy of pre-tender building cost estimate", *Management and Innovation for a Sustainable Built Environment*, Amsterdam, The Netherlands, 20 – 23 June 2011,

Al-Tabtabai, H., et.al., (1999) "Preliminary cost estimation of highway construction using neural networks, *Construction Engineering*, Volume (41): 19-24.

Arafa, M. and Alqedra, M., (2011) "Early stage cost estimation of building construction projects using artificial neural networks", *Journal of Artificial Intelligence*, Volume 4(1): 63-75.

Bode, J. (1998) "Neural Network for Cost Estimation", *Cost Engineering*, Volume (40)1: 25-30.

Bode, J. (2000) "Neural Network for Cost Estimation: Simulations and Pilot Application", *International Journal of Production Research*, Volume (38) 6:1231-1254.

Camargo, et.al., (2003) "Application of the parametric cost estimation in the textile supply chain, Journal of Textile and Apparel", *Technology and Management*, Volume (3): Issue 1.

Chan, S. and Park, M., "Project cost estimation using principal component regression", *Construction Management and Economics*, Volume (23): 295-304.

Cunningham, P., (1998) CBR: Strengths and Weaknesses, 11th International Conference on

Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, eds A. P. del Pobil, J. Mira & M. Ali, Lecture Notes in Artificial Intelligence 1416, Volume (2): 517-523,

Emsley, M. et.al., (2002) "Data modelling and the application of neural network approach to the prediction of total construction cost", *Construction Management and Economic*, Volume (2)0: 465-472.

Fang, C. and Froese, T. (1999) "Cost estimation of high performance concrete (HPC) high-rise commercial buildings by neural networks", Durability of Building Materials and Components 8. (1999) Edited by M.A. Lacasse and D.J. Vanier. Institute for Research in Construction, Ottawa ON, K1A 0R6, Canada, pp. 2476-2486.

Ganiyu, B. and Zubairu, I. (2010) "Project cost prediction model using principal component regression for public building projects in Nigeria", *Journal of Building Performance*, Volume (1): Issue 1.

Garza, J. and Rouhana, K. (1995) "Neural Network versus Parameter-based application", *Cost Engineering*, Volume (37)2:14-18.

Gould, F. and Joyce, N. (2000) Construction Project Management. Prentice Hall, Englewood Cliffs, NJ, 2000.

Gunaydin, H. and Dogan, S., (2004) "A neural network approach for early cost estimation of structural systems of building", *International Journal of Project Management*, Volume (22): 595-602.

Han, K.H., et.al., (2008) Cost Estimation Methodology Using Database Layer in Construction Projects, The 25th International Symposium on Automation and Robotics in Construction, June 26-29, 2008.

Kim, D.Y., (2008) Two-staged early cost estimation for highway construction projects, The 25th International Symposium on Automation and Robotics in Construction, June 26-29, 2008.

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*. Volume (39): 1235-1242.

Manprasert, S., (1998), A Study on Building Construction Cost Estimation Using Work Quantity Model Approach, Thesis, Faculty of Engineering, Chulalongkorn University, Bangkok, Thailand.

Mohamed, A. and Celik, T. (2002) "Knowledge-based system for alternative design, cost estimating and scheduling", *Knowledge-based Systems*, Volume (15): 177-188.

Montgomery, D. and Peck, E. (1982) Introduction to Linear Regression Analysis, John Wiley and Sons, New York.

Neural Network Toolbox in MATLAB 7.10.0 (R2010a)

Olotuah, A. (2002) Recourse to Earth for Low-Cost Housing in Nigeria. Building and Environment, 37, 123–129, 2002.

Oreta, A. (2012) Artificial neural networks: Multi-layer feedforward networks using back propagation, CIV578D Lecture Notes, De La Salle University – Manila.

Shtub, A. and Versanob, R., (1999) "Estimating the cost of steel pipe bending, a comparison between neural networks and regression analysis", *International Journal for Production Economics*, Volume (62), Issue 3, 20 September 1999, 201–207.

Smith, A.E. and Mason, A. K. (1997) "Cost Estimation Modeling: Regression versus Neural Network", *Engineering Economist*, Volume (42)2: 137-1361.

Squeira, I., (1999) Neural network based cost estimating, Master's Thesis, Department of Building, Civil and Environmental Engineering, Concordia University, Montreal, Quebec, Canada.

Tewari, D. and Singh, K., (1996) Principles of Microeconomics, New Age International Publications, New Dehli, page 342.

Turochy, et.al., (2001) Highway project cost estimating methods used in the planning stage of project development, *Technical Assistance*

*Report*, Virginia Transportation Research Council.

Yang, K., et.al, (2008) "Neural network modelling of RC deep beam shear strength", *Structural Building*, Volume 161:29-39.