

Artificial Neural Network Models: Conceptualized and Improved Method of Contingency Assessment for Nigerian Building Projects

Momoh Ohiomah Oboirien

Department of Building, School of Environmental Sciences,

Modibbo Adama University, Yola, Adamawa State

Email: oboirien.momoh@mautech.edu.ng

Abstract:

Precision in estimating the contingencies in project parameters has been difficult in Nigeria despite the numerous available computational techniques. This study aims at exploring the predictive ability of artificial neural network with a view of designing prediction models and ascertaining efficiencies. Secondary data from a recently completed research based on projects in the North Eastern Nigeria were analyzed. Their surveyed construction cost ranged between ₦103.80 million and ₦303 billion, while delivery periods were between 20 and 192 weeks. Others are high cost and duration drivers, which formed the artificial neural network (ANN) models input variables, 210 cost and 199-time data sets formed the output variables. Mean contingencies of 6.46% and 7.06% for cost and duration were established from the developed models. The study concludes that the developed ANN forecasting models can aid the construction contractor in estimating project contingency of higher precisions. It, therefore, recommends the conversion of the developed models into Dashboards for contractors' use.

Keywords: Artificial neural network, contingency, cost and time modelling, estimating, North-East Nigeria

1 Introduction

Determining the comprehensive budget and duration of proposed construction projects have been difficult. This is because of the difficulty in ascertaining the contingencies that are inevitable while constructions are in progress. Contingencies are extra funds and time included within the budgets and construction programmes so that total finances and schedules can be committed. In other words, contingencies are provisions for absorbing the increments in funds and schedules that are unavoidable in managing the risk of overruns on project goals (ACE, 2010). Accurate estimations of contingencies are, therefore, of critical importance to construction project success. Such reserves in funds and time can significantly impact when the estimates are higher than the optimum, thereby encouraging untidy project cost and schedule control. Invariably, the completed buildings become costly and unprofitable and deprive the company of monies that could have been invested elsewhere. If the contingencies are underestimated and rigid, they could set unrealistic financial and duration issues, and the consequences are usually unsatisfactory performance. Questions this study intends to answers are how effective have been the contingency sums hitherto included in the building

construction projects vis-a-vice the methods used in the derivations? Can the current methods previously used in the provisions of contingency sums in Nigeria be described as satisfactory? Can a more scientific method of contingency computation for building construction projects be explored?

This study aims at exploring the forecasting power of ANN with a view of designing building construction project cost and time prediction models and ascertaining the efficiencies. The objectives are to (i) investigate the current estimating techniques used for deriving the contingency sums for building project budget and schedule (ii) develop ANN models for project cost and schedule estimation (iii) establish building project budget and schedule contingencies from the developed ANN models. The scope is limited to 246 completed government-owned buildings in the North-East geopolitical zone of Nigeria. Results of this study shall have a significant impact on future building construction project economic management.

2 Contingency Sums Estimating Techniques

A set of techniques for construction resources estimation exist for computing the budget and time contingencies. These are Analytical Hierarchy Process, Fuzzy Sets, Influence Diagrams, Individual Risks – Expected Value, Factor Rating, Methods of Moments, Monte Carlo Simulation, Range Estimating, Regression Analysis, Theory of constraints, Traditional (Customary) Percentage, and recently ANN. The customary percentage method is the most used; however, Monte Carlo Simulation, Regression Analysis have become famous while ANN modelling technique is increasingly being embraced in the advanced economies. The customary method of contingency sum determination has continued among estimators for the past years. It is a general percentage increment of the basic quantity of cost or schedule, usually recalled instinctively or obtained from records and knowledge from past projects. For instance, an addition of about 10 – 20 per cent of the basic, to determine a working estimate. This is common to firms in the construction industry.

According to Oтали & Odesola (2014), the customary method is arbitrary, subjective, unreliable, grossly inadequate, and it is not easy to defend or justify as it is unscientific and helps in over budgeting. This research investigates Ameh *et al.*'s (2010) suggestion of a thing that should evolve with scientific management principles. Its results hopefully will replace the general and unscientifically related addition to the basic sum.

Recently, the weakness of the general addition approach has led to research into more robust methods of contingency funds and extra time determination as artificial neural network. The use of ANNs for project parameters estimation has grown such that the models can now be used to predict project cost and time overruns that are needed as precautions right from project onset. When adequately computed, they become management tools and contingency approaches of higher precision than current methods.

2.1 The ANN Models

Some construction industry researchers had ventured into ANN with recorded success because of the numerous advantages. These include (i) ability to appropriate, expressing, and mapping facts involving two or more variable quantities to an additional set of details, (ii) information processing in a pattern like the human brain, (iii) power of deriving sense from sophisticated, random, and inexact statistics, (iv) detecting inclinations that are compounded and certainly not done by alternative problem-solving devices (v) it is used for projections where new situations of interest arise, and (vi) ability to answer hypothetical questions. A good number of problems that have defied rule-based programming in different fields have been solved with ANN, because the model works on a training and learning concept. Like the stochastic relationships between the multivariate construction cost and time drivers and the associated impacts on the initial targets, the desired models should be most suitable.

In the network architecture, the value of all weighted inputs impels the degree of the level of inducement that is further altered by a triggering function to produce the output signal, summarized in an output represented by 0. The output is denoted by $f(\sum_i W_i X_i)$. X_i and W_i are the i th input and matching weight. The stimulation can be a threshold function, or a smooth function like a sigmoid or hyperbolic tangent function and is represented by f . Neural network representations in machine learning or artificial intelligence (AI) are commonly referred to as ANNs; which are in essence simple computative equations defining a function $f: X \rightarrow Y$ or a spread on X or both X and Y .

2.2 Contingency computations for construction projects using ANN

Chen & Hartman (2000) is one of the past studies that utilized ANN to forecast the final amount of a concluded construction project. The authors found that 75% of the predicted final figure aligned with the real deviations, pointing to the fact that variabilities indeed happened as predicted by the network. This meant that if the part left out of 100 ($100 - 75$) i.e., 25 % were to be provided as extra fund as contingent sum, there could have been nothing left to call overrun or underrun. Oboirien (2019) computed sums complementary to the efficiencies of twelve ANN construction parameters estimating models and found an average of 10.21% herein called contingency sum.

2.3 The research concept

The persistent occurrence of budget and schedule slippages in building work globally indicates that more research effort is required if the outcomes are to be kept close to the set objectives. The evidence is found in the works of Flyvbjerg (2005), Edwards & Kaeding (2015), Idiako *et al.* (2015), Amusan *et al.* (2017), where projects overran their initial cost and time parameters by over 50 to 100%. In Nigeria, the slippages are up to 500%. For example, Oraegbune (2008) found 683.33% time overrun on NIPOST Headquarter buildings, 400% cost overrun on Upper Benue River Basin Development Authority Dam, all in Yola, Adamawa State capital. Since, estimators normally allow for budget and programme risk mitigations (contingencies) at planning stage, the current high variability between the initial and final project outcomes supports the argument that an innovative and more robust technique over the orthodox and popular ones is now needed in the industry.

Mean contingency allowance 10.21% common to projects estimated with ANN models is very well lower than the overruns recorded globally, ranging between 36.02 to 138%. Obviously, if all construction project parameters were forecast with ANN, the current unbearable slippages in the global construction industry would obviously be minimal and manageable. Currently, only about 7 % of research on budget and schedule assessments that applied ANN forecasting technique are based on the African context (Oboirien, 2019). More effort is still required in conceptualizing solutions to the current high construction cost and time slippages via a more precise contingency estimation method in developing economies like Nigeria. This research is therefore focusing on addressing the knowledge space (gap) articulated in the preceding section.

3 Research Methodology

The study centre was North Eastern Nigeria which comprised six states but excluded two: Borno and Yobe, due to the intense insurgency activities of an Islamic sect (Boko Haram). The study sample was 246 public building projects spread across Adamawa North, Adamawa South, Bauchi State, Gombe and Taraba States. Adamawa was divided into two to ensure a reasonable project spread. While the construction cost ranged between ₦303 billion and ₦103.80 million, the shortest construction duration of the project studied was 20 weeks; the longest was 192 weeks.

3.1 The study's ANN models development steps

Five steps were involved: data gathering, examinations and presentation, followed by network layout definition, the learning step, training and testing of the model and lastly, establishing the models' efficiency levels (validations).

3.2 Determining the ANN input and output variables

Oboirien (2019) identified the main driving factors on 210 and 199 completed buildings in the study area that formed the ANN models' input (influence) and output (impact) variables. The cost and duration factors' influence were computed from research participants' responses from the administered questionnaire. The performance impacts (model outputs Z_c and Z_t) on the project and planned durations were calculated with the following equations (1) and (2). The values were fed into the ANN software.

$$\text{Output variables } Z_c = \frac{\text{Final cost} - \text{Initial contract sum}}{\text{initial contract sum}} \quad (1)$$

$$\text{Output variables } Z_t = \frac{\text{Actual construction duration} - \text{Estimated construction duration}}{\text{Estimated construction duration}} \quad (2)$$

3.4 Division of data into training, learning and validation sets

The cleansed input and output data sets were divided into two in a ratio of 80% to 20% (Gunaydin & Dogan, 2004). The divided data are 209 for cost and 198 for the time. Though there are not yet acceptable generalized rules for deciding the size of suitable training data; Setyawati *et al.* (2002) advised for training samples that should cover all sections of the available statistics.

3.5 Data mining for pattern learning

The larger of the two divided data sets was used for detecting inhibited patterns in the collected data (data mining). In this step, any patterns or relationships in the variables found to be continuously uniform are used on new data sets for purposes of checking or proving acceptability. This step aims to search for relationships that have predictive potentials for improving an existing system. For instance, a mechanism for forecasting project parameters in which differences between the initial and outcomes can be consistent in various data sets, which are acceptable differentials (contingency) that construction project planners need the awareness and factored-in at the pre-contract stage. This dimension is yet to gain popularity in construction research, but Yang *et al.* (2003) found its use in construction equipment productivity estimating. Other studies that have benefited from data mining are Yu and Lin (2006), and Cheng *et al.* (2012). The studies are respectively in construction management knowledge refinement, workplace accidents and concrete strength forecasting.

This is the model teaching stage that needs the sourced data prepared and acquire the model learning principles. Backpropagation (backward propagation of errors) is the intrinsic nature of neural net training. It is the method of fine-tuning the weights of a neural net based on the prevailing rate of error in the preceding epoch (i.e., iteration). Backpropagation is an accepted mode of nurturing ANNs. It aids in computing the slope of a loss function regarding all the weights in the network. Appropriate tuning of the weights allows for reducing error rates and makes the network well-grounded since it increases its judgement. It is classified as supervised machine learning because it requires a known output for each input value. It computes the loss function gradient to determine the unknown relationship in a repeated data feeding.

Model training is stopped when a stated number of iterations is reached. Performance measures are applied to pick the best network configuration, the training rules, and the criteria. These are Root Mean Squared Error (RMSE) given in equation (3).

$$\text{RMSE is expressed as: } \sqrt{\frac{\sum_{i=1}^n (t_i - O_i)^2}{n}} \quad (3)$$

Number of projects in the sample is represented by n , t_i actual impacts computed from survey data, and O_i is model's predicted impacts.

4 Findings and Discussions

4.1 Details of research participants

246 respondents from the study centre each supplied data on a project. Details of professions, membership of professional associations, stakeholder category and years of industrial experience are shown in Figures, 1, 2, 3 and 4. Participants were built-environment professionals i.e., registered builders (45.53%), while mechanical, electrical, civil, and structural engineers combined (25%). Others were architects (21%), and quantity surveyors (8%) depicted in Figure 1. A breakdown of participants' corporate membership of professional association indicates that all were in the full membership grades, except the builders, quantity surveyors, and engineers with a small percentage on graduate membership grades (Figure 2).

A large portion was in the employment of main and subcontractors, while others worked in corporate clients and consultancy firms (Figure 3). In terms of years of experience, there were five classes. Between 11-15 years fell into the modal class of 19 %. The lowest and largest classes of 16-20 and 6-10 years are 9% and 33 %, respectively (Figure 4). The supplied data are, therefore, adjured fit for use in the analysis.

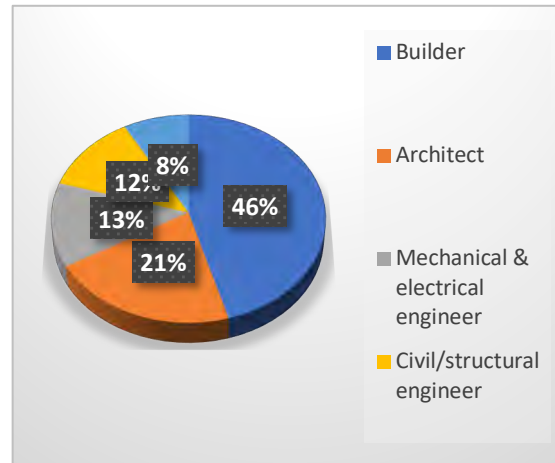


Figure 1. Respondents' professional background details

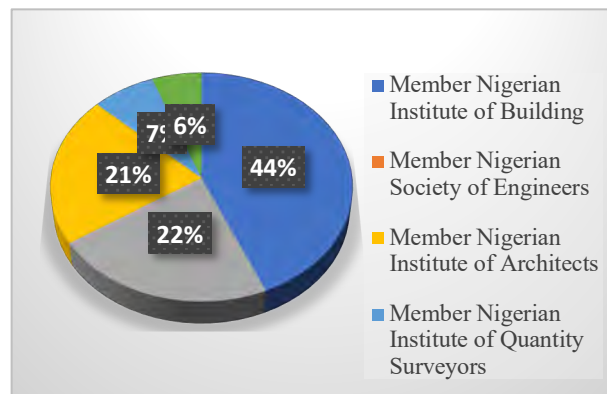


Figure 2. Respondents' membership of professional association

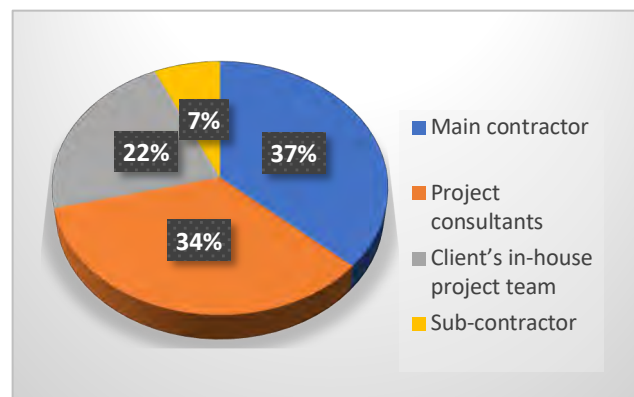


Figure 3. Respondents' stakeholder category

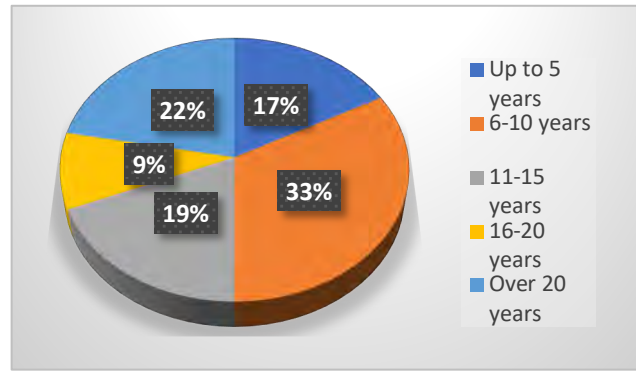


Figure 4. Respondents' professional experience

4.2 The ANN building construction project cost and time contingency estimation models

Pairs of the processed independent and dependent data sets were trained using Smart Lab, an ANN software. To understand the pattern of association between the inputs and output data, 294 cases were reduced to 209 datasets, due to incorrect filling of questionnaire by the research participants and a sieve from outlying data. These were used in the analysis. Using the 80: 20 % data ratio as stated earlier, 168 datasets which were 80% were used in training the cost influence and impact prediction network. The other 41 datasets were used in confirming if the outputs of model are acceptable (validation).

Model training began with the data divided into 70:30 ratio, and sample adjustments till it had a good performance at 80:20 division. Trials' RMSE of the trained model was computed using equation (3). This good performance is an indication that the interconnection is now reasonably knowledgeable about the association between the inputs and outputs and could effectively generalize from fresh data sets. The iterations resulted in 18 hidden nodes of two layers. Normally enough processing elements are based on hidden layer trials as there are yet to be the authoritative rule for determination. The model after many iterations, went through a gradual transformation process in a backpropagation that resulted to 9-9-9-1 layout. 0.75, with 5288 training cycles was found as best network from the series of trials, with the set error in training reduced to 0.001. The stabilized and final model becomes a prediction model for construction cost forecasting. The subtraction of the model efficiency determined using the reserved validation data, from one hundred gave the desired cost contingency estimate, by the ANN construction cost forecasting model.

In similar trials, network model training for the construction factors' influence and impact on planned duration stabilised at the 80:20 data division. 159 datasets represented 80% of the survey figures were used for the duration impact network training, while 39 raw data were reserved for use in confirming the model's level of training. The resultant network configuration was 10-6-6-1. In this case, the impact training cycles (epochs) read 1534 while the learning error was 0.80. Training inaccuracy was configured to gradually decrease and get as small as one-thousandth (0.001).

4.3 ANN models validations and construction project cost and time contingencies

The validation tests conducted on the developed network models were comparable between the predicted and real values, using the reserved raw information. The computed RMSE statistic enabled validation of the trained models. The forecasting efficacy of the developed cost impact network is 93.54%. This implies a contingency of 6.46% when 93.54 is subtracted from 100.

In a similar process, the models' development process evaluation conducted were considerations of similarities between the predicted and the true impacts from the designed duration model and those computed from the reserved sourced data. Again, the statistical verification technique used here is RMSE. The 39 datasets representing 20% of the survey data reserved for testing the forecasting precision of the duration network were used. Input variables that influence the driving factors, and output variables which were the duration impacts or extra times over the planned were entered into the model. The ANN network was asked to forecast the duration impacts (ratios of the difference between estimated and actual construction times on the originals). The RMSE statistic of the developed duration impact forecasting model represents the required construction programme contingency. This was determined when the validated construction duration impact estimating model efficiency is subtracted from a hundred ($100 - 92.94$). Therefore, construction programme contingency found in this study is 7.06%. The study results compare favourably with related research outside Nigeria. For instance, Gunaydin and Dogan (2004) made use of eight design criteria to estimate the price per unit area of a reinforced concrete structural complex in low-rise private homes; the network produced a budget precision of 93%. The remaining 7% serves as a needed contingent sum. Similarly, the work of Aibinu, Dissanayaka, Chan, and Thangaraj (2015) on cost estimation for electric and power elements during building design, using a neural network approach, resulted in forecasted contingency sums of 6.4, 4.5 and 4.5% for cable pathway, light wiring, and power wiring, respectively.

Since Chen and Hartman (2000) revealed the predicted values same as actual project values, it can be concluded that ANN models for construction project parameter estimating have higher prediction accuracies than the traditional methods still popular among Nigerian estimators. The network prediction models are, therefore, better. Since, construction project planners normally allow for budget and programme risks mitigation from onset using at least one convention technique. The current industrial status in Nigeria (where budget and schedule slippages in multiples of hundreds are still prevalent) beckons on the researcher to explore further into improved construction resources estimation technique like the ANN prediction models, designed in this study.

5 Conclusions and Recommendations

Ascertaining the inevitable project contingencies with precision has been difficult, and not much has been achieved in Nigeria with current methods. The study explored the predictive ability of ANN with a view of designing building project budget and construction duration prediction models as well as ascertaining their efficiencies. The objectives were, (i) investigate the estimating techniques used for computing project cost and time contingencies (ii) develop

ANN estimating models for construction project cost and time (iii) establish building construction project budget and schedule contingencies from the developed ANN models.

Twelve forecasting methods for budget and activities programme contingencies were found in the existing literature. They are Analytical Hierarchy Process, Fuzzy Sets, Influence Diagrams, Individual Risks – Expected Value, Factor Rating, Methods of Moments, Monte Carlo Simulation, Range Estimating, Regression Analysis, Theory of constraints, and Traditional Percentage. Their use, however, are yet to meet the desired industrial satisfaction in terms of zeroing out project overruns via a highly efficient contingency forecasting technique.

This research developed ANN cost and time prediction models that validated efficiencies 93.54% and 92.94%. These translate to 6.46% and 7.06% cost, and duration percentage complements, herein referred as estimated contingencies using the data sourced from government-owned buildings in Nigeria's North East geopolitical zone. The implications are ANN model-based project budget and duration predicted, can assist the constructor in estimating contingencies of higher precision than the methods hitherto used in the industry. It, therefore, recommends the conversion of the developed models into Dashboards for use by building project designers and managers.

6 References

- Aibinu, A. A., Dissanayaka, D., Chan, T., & Thangaraj, R. (2015). Cost estimation for electric and power elements during building design: A neural network approach. *Engineering Construction and Architectural Management*, 22(2), 190-213.
- Ameh, O. J., Soyingbe, A. D., & Odusami, K. T. (2010). Significant factors causing cost overruns in telecommunication projects in Nigeria. *Journal of Construction in Developing Countries*, 15(2), 49 – 67.
- Amusan, L., Dolapo, D., & Joshua, O. (2017). Cost and time performance information of building projects in developing economy. *International Journal of Mechanical Engineering and Technology* 8(10), 918–927.
- Association for the Advancement of Cost Engineering (AACE) International (2010). Cost engineering terminology. *AACE Recommended Practice No 10S-90. TCM Framework: General Reference*.
- Chen, D., & Hartman, F. T. (2000). A neural network approach to risk assessment and contingency allocation, *AACE Transactions*, 24-27.
- Cheng, C. W., Leu, S. S., Cheng, Y. M., Wu, T. C., & Lin, C. (2012). Applying data mining techniques to explore factors contributing to occupational injuries in Taiwan's construction industry. *Accident Analysis & Prevention*, 48(0), 214-222.
- Edwards, C., & Kaeding, N. (2015). Federal government cost overruns. *Cato Institute Tax and Budget Bulletin No 72*. Retrieved from www.downsizinggovernment.org/government-cost-overruns. Assessed 02/04/2016.
- Flyvbjerg, B. (2005). Machiavellian megaprojects. *Antipode*, 37(1), 18.
- Gunaydin, H. M., & Dogan, S. Z. (2004). A neural network approach for early estimation of structural systems of buildings. *International Journal of Project Management*, 22, 595-602.

- Idiake, J. E., Abdullateef, A. S., & Abdulganiyu, A. O. (2015). A study of time and cost relationship of private building projects in Abuja. *International Journal of Construction Engineering and Management*, 4(1): 26-34.
- Oboirien, M. O. (2019). *Modelling cost and time performance of public building projects in a terror impacted area of Nigeria*. Thesis Presented for the Degree of Doctor of Philosophy Department of Construction Economics & Management. Faculty of Engineering and the Built Environment, University of Cape Town.
- Otali, M., & Odesola, I. A. (2014). Effectiveness evaluation of contingency sum as a risk management tool for construction projects in Niger Delta, Nigeria. *Ethiopian Journal of Environmental Studies and Management* 7(6), 588-598. Retrieved from <http://dx.doi.org/10.431/ejesm.v7i6.1>. Assessed 09/09/2015.
- Setyawati, B. R., Sahirmann, S., & Creese, R. C. (2002). Neural networks for cost estimation. *AACE International Transactions*.
- Yang, J. B., Edwards, D. J., & Love, P. E. D. (2003). A computational intelligent fuzzy model approach for excavator cycle time simulation. *Automation in Construction*, 12(6), 725-735.
- Yu, W. D., & Lin, H. W. (2006). A VaFALCON neuro-fuzzy system for mining of incomplete construction databases. *Automation in Construction*, 15(1), 20-32.