

MODEL FOR ESTIMATING CONSTRUCTION COSTS FOR LOW-RISE RESIDENTIAL BUILDINGS

ADEL ALSHIBANI¹, OTHMAN ASHAMRANI² & MESSAM SHAAWAT²

¹Department of Architectural Engineering, KFUPM, Kingdom of Saudi Arabia

²Department of Building Engineering, Imam Abdulrahman Bin Faisal University, Kingdom of Saudi Arabia

ABSTRACT

This paper introduces a multi-regression model for estimating construction cost of various structure and envelope types of low-rise residential buildings in Canada. The model is capable of predicting construction cost per square feet for six envelope and structure alternatives in various combinations. The structure types were of wood and steel, and the envelope systems included wood, veneer brick and concrete cavity wall. The predictor variables were building envelope type, area, number of story and story height. The model was developed in four main steps: literature review of the existing methods, real construction cost data collection of completed projects, preliminary diagnostics over data quality, generation and verification of the model. The developed model was successfully tested and validated with real-time data. This model can provide reliable conceptual cost estimate of low-rise residential buildings at the early design stage with reasonable accuracy. It helps owners for budget allocation and/or conducting economic analysis with less effort and complexity.

Keywords: construction costs, low-rise residential building, regression model, normal distribution, residual analysis.

1 INTRODUCTION

Estimating construction cost at the planning stage where there is no enough information available is a difficult task, owing to the uncertainties associated with future cost. Researchers and specialists have identified the uncertainties associated with the estimation of construction cost and the need to enhance the performance of prediction models [1]. Substantial efforts have been put to address this issue and a lot of conceptual cost prediction models are currently available in practice, based on techniques such as Genetic Algorithm(GA), Probabilistic Cost Estimation, Case-Based Reasoning (CBR), Regression Analysis, Fuzzy Logic (FL), Neural Network (NN), and so on.

The relative merits and demerits of these techniques were analyzed by experts, which are well documented [2], [3]. However, a review of the updated literature related to the current study is presented here. Regarding the use of regression analysis, Li et al [4] proposed step-wise liner regression models for office buildings in Hong Kong, while a multivariate regression model named as estimate score procedure was developed by Trost and Oberlende [5]. In a similar study, linear regression models were developed for the prediction of construction cost of United Kingdom's buildings [6], based on 286 different sets of real data. Application of Neural Networks (NNs), Fuzzy Logic (FL) and Genetic Algorithm (GA) for construction cost prediction has attracted the researchers and practitioners, and the literature is abundant in this area. Siqueira [7] applied NNs for cost estimation of low-rise prefabricated structural steel buildings in Canada. The data were collected from 75 completed building projects over a 3-month period. Similar study was reported from Turkey [8], which used data from 30 projects to train and test the NN model developed for cost prediction of 4 to 8 story residential buildings

Kim et al. [9] incorporated GA in their back-propagation network (BPN) model to improve construction cost estimation accuracy. For the training and assessment of the model, the construction data of 530 residential buildings in Korea was taken between the time



periods of 1997 to 2000. Yu et al. [10] developed Web-based Intelligent Cost Estimator (WICE) model that incorporated the features of data mining, neuro-fuzzy system and WWW. The proposed model was claimed to provide an adequate, globally accessible and reliable decision maker tool in real time that could also provide efficient feedback. A subsequent study [11] proposed Evolutionary Fuzzy Neural Inference Model (EFNIM), which incorporated GA, FL and NNs features. The EFNIM was then combined with WWW and historical data to form Evolutionary Web-based Conceptual Cost Estimators (EWCCE) which provided two kinds of estimators for conceptual construction cost. The Artificial Neural Network (ANN)-based evolutionary fuzzy hybrid neural network (EFHNN) developed by Cheng et al. [12] was claimed to be effective for precise cost estimation of construction projects during their initial stages. Few other recent NN-based models included those reported by Juszczak [13], Bala et al. [1] and Aibinu et al. [14].

In the CBR model, new problems are solved by providing the solutions of already known earlier and similar problems [15]. Many works were reported on developing models based on CBR. For instance, An et al. [15] proposed a CBR model based on analytic hierarchy process (AHP), which included all the processes of cost estimation; few similar models were developed by Koo et al. [16], Hong et al. [17] and Ji et al. [18]. An advanced CBR model was presented by Koo et al. [19], containing multi-family housing projects with 101 cases; the model incorporated optimization process, ANN and multiple regression analysis (MRA), using GA. The proposed user-friendly model was developed by using Visual Basic that was connected with Microsoft-Excel data base.

Attempts were also reported on developing new and hybrid prediction models. The On-Line Analytical Processing (OLAP) environment introduced by Moon et al. [20], the Principal Item Ratios Estimating Method (PIREM) proposed by Yu [21], and the bootstrap approach presented by Sonmez [22] are good examples of the new approaches. Forecasting models of initial costs are developed for school and college buildings in North America by Alshamrani [23], [24]. However, there is enough scope to develop models for estimating construction cost of low-rise residential buildings, particularly for comparing the costs of different structure and envelope types and selecting the economically viable alternatives for the building.

This paper demonstrates a regression model for the prediction of construction costs of low-rise residential buildings in Canada. The construction costs are estimated for six structural and envelope alternatives with floor height, specific area and floor numbers. The real cost data of completed projects was used to develop the model.

2 METHODOLOGY

Real cost data of completed projects was utilized in this paper to develop the construction cost prediction model. Various input parameters were identified and described to estimate the construction costs. These parameters were envelope type, floor height, building area and structure type. The data used as input to develop the model included:

Building area: 7000, 10000, 20000, 30000 and 40000 ft²

- Floor No.: 1, 2, and 3 floors.
- Height of floor: 10, 11, 12, and 13 feet.
- Type of Structure: Steel frame (S) and wood frame (W).
- Type of Envelope: wood sides on wooden studs (W), concrete brick backed up with concrete block (C), and insulated veneer brick (V).
- Location: National average for Canadian cities.
- Year of construction: collected data for different years.



2.1 Construction cost breakdown

Construction costs for new apartment buildings were estimated after defining the parameters by the model. Construction cost included the following:

- Architecture fee: supervision, drawing and design.
- Contractor fees: profits, contingency, overhead and general conditions.
- Equipment and furnishings: HVAC, institutional and other equipment.
- Interiors: ceiling finishes, floor finishes, wall finishes, stair construction, fittings, interior doors and partitions.
- Services: electrical systems, security and communications, branch wiring and lighting, electrical distribution and services, standpipes, sprinklers, cooling systems, energy supply, rain water drainage, distribution of domestic water, plumbing fixtures, lifts and elevators.
- Super structure (Shell): roof construction, roof openings, roof coverings floor construction, doors, windows and exterior walls.
- Substructure: walls, excavation of basement, slab on grade and foundations.
- Subtotal cost is calculated by adding component cost breakdown after their estimation. The subtotal cost is then added to fees of architectures and contractors.

2.2 Preparation of data for modeling

A real construction cost data of completed projects was collected and used as input parameters in the form of independent variables. These parameters were: number of buildings, number and height of floor, envelope type, year of construction, location and structure type. All the parameters listed in Fig. 3 proved meaningful effect on construction cost. All of these factors and parameters were studied in this paper to identify their relationship with construction cost which is a dependent factor. The starting cost calculated through field study consisted of about 360 data points out of which 80% (300 points) were used to build the prediction model for the construction cost of low-rise residential buildings, while 20% (60 points) were picked randomly to validate the model.

The primary objective of the present model is to discover the relationship between response variables and predictors. To state the relationship and build the prediction model for every envelope and structure type, multiple linear regression technique was used. As illustrated in Fig. 1, the development of regression model involved four main stages: collection of actual construction cost data of completed projects, preliminary data quality diagnostics, process for model generation and validation of model. Collecting real construction cost data involved gathering 360 data sets of completed projects throughout the country, considering different cities and different years (age) of building. Preliminary data investigation contained two things: identifying any data interaction and correlation, and carrying out the analysis for the best subset regression. Process for model development consisted of four steps: regression model generation, examination of elementary factors, residual study and validation model selection.

2.3 Data collection

Poor documentation of previous construction costs of residential buildings made finding the required data with certain specifications to build the model a difficult task. Therefore, the required data was collected through field survey, personal communications and online



questionnaire. The collected data consisted of the aforementioned independent variables (input data) for the developed regression model and one parameter representing construction cost (output). The data was filtered, and incomplete data points were removed before analysis. Ultimately, 360 data points were used to build the model to predict the construction cost of residential buildings. The data was organized in Microsoft Excel worksheets to allow an easy application of the regression software used to build the model.

2.4 Preliminary data diagnostics

2.4.1 a) Stating interactions and correlations

The first step in preliminary data analysis was to identify any possible interactions of the developed model's predictor variables or to state any multi co-linearity. Correlation was determined through the simulation of matrix scatter plot between response factor and predictor variables. The representation of scatter plot is much essential to identify the correlation and data linearity among response variables and predictors and predictor variables

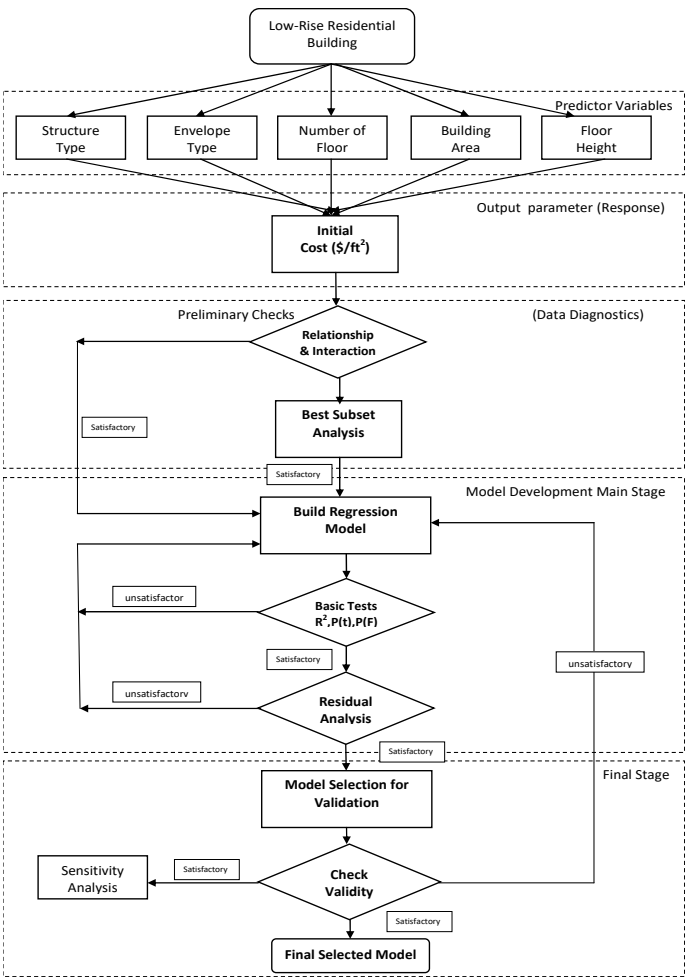


Figure 1: Regression model development process.

2.4.2 b) Best analysis for subset

This analysis determines least variation and error values as well as predictors' best possible combination related to the highest values of R^2 (adjusted) and R^2 . Therefore, the regression model that can be best-fit was developed through regression analysis of best subset along with certain variables.

3 GENERATING OF REGRESSION MODEL

The generation of best data regression model was performed after identifying the correlation based on best data set of the collected real data. Through this model, building owners can forecast the construction cost of new low-rise residential building having different envelope and structure types. This regression model is quite user-friendly and has best-adjusted data on hands. The calculated data was saved in Excel sheets to achieve flexible analysis of spreadsheet. Development of regression model was performed through a statistical software package named as Minitab®. Regression analysis was performed by distorted variables installation in Minitab®. Minitab® output contains regression equations that have regression coefficients " β_k " for examined data.

3.1 Preliminary assessment for model adequacy

The regression model's preliminary assessment consisted of multiple determinations coefficient (R^2), a regression relation test (F) and a test (t) related to coefficient " β_k " of regression parameter. The value of (R^2) calculates the data fitting in correlation to "construction costs" (response variables) or the variance of predictor variables. Several model predictors were calculated by R^2 (adjusted). Both of these values are responsible to fit the data in a fine manner.

In the regression relation test, a hypothesis test was implemented to measure P (F) for entire model. All the regression coefficients $\beta_0 = \beta_1 = \beta_{p-1} = 0$ were considered to be zero under the supposition of null hypothesis (H_0) while the supposition of alternate hypothesis (H_a) addressed that all of the coefficients were not considered to be zero. Null hypothesis was rejected when p-value (statistical significance) was 0.00. Estimated model was proved to be significant at 0.05 value at α -level due to this hypothesis, which shows that at least one coefficient has some value instead of zero in the regression model.

The third test is related to the correspondence of all predictors to response variables. Regression coefficient efficacy and $\beta_0, \beta_1 \dots \beta_{p-1}$ values were determined through individual "t-tests". In the case of β_0 , the alternative hypothesis (H_a) of the t-test considered that $\beta_0 \neq 0$ while null hypothesis (H_0) considered that $\beta_0 = 0$.

The next step was the analysis of residuals and their impressions after the satisfactory diagnostics of bases and coefficients. The verification of linearity correlation suppositions was carried out through normality error check. The graphs constructed in the developed model represented normal frequency and probability through which residual analysis was performed.

3.2 Validity assessment of regression model

The first step for validation was related to the comparison of anticipated values with the substantial observation in which data is verified for every developed model. The validation was carried out through points of 20% excluded data which was then plotted to make a comparison among in-hand observed data and model of prediction. The mathematical model

used for validation was adopted from Zayed and Halpin [25], which calculate validity percent and average invalidity to validate the data as follows:

$$AIP = \frac{\sum_{i=1}^n \left| 1 - \left(\frac{E_i}{C_i} \right) \right|}{n} \quad (1)$$

and

$$AVP = 1 - AIP, \quad (2)$$

where AIP, AVP, E_i , C_i and n denote average invalidity percent, average validity percent, Predicted Value, Actual Value and number of observations respectively; the value of AIP differs from 0 to 1.

4 RESULTS AND DISCUSSION

The prediction of square foot construction cost of residential buildings is performed through Multi regression model. This model also predicts specified envelope and structure type with the consideration of correlated predictor variables. These variables include: floor height ranging from 10 to 13 feet, floor number ranging from 1 to 3 and area of building in square foot.

Linearity of data is tested in the first place by identifying the relationship in transformed Y correlation matrix and acquired scatter plot matrix. The plots are shown in Fig. 5. Data is evenly distributed and constant in the entire graph that also avoids making any patterns. These plots demonstrate that the predictor variables are almost linear related to the response variables, and hence the plots are acknowledged to be satisfactory

4.1 Best subset

Fig. 2 shows that different regression models are generated in each line by the subset regression analysis output. The smallest values of standard deviation (S) and C_p are 4.46 and 6.0 respectively, where the peak values for R^2 and R^2 (adjusted) are 85.3% and 85.0% respectively in WC regression model. Results of best subset analysis illustrate that all the predictors are significant which must be a part of regression model development.

Developed Regression Model:

$$CC = 166.68 + 3.948N - 0.000753A + 1.203H + 5.152S - 8.066E, \quad (3)$$

where:

CC: Predicted Construction costs,

N: Number of floors which ranges between 1–3 floors

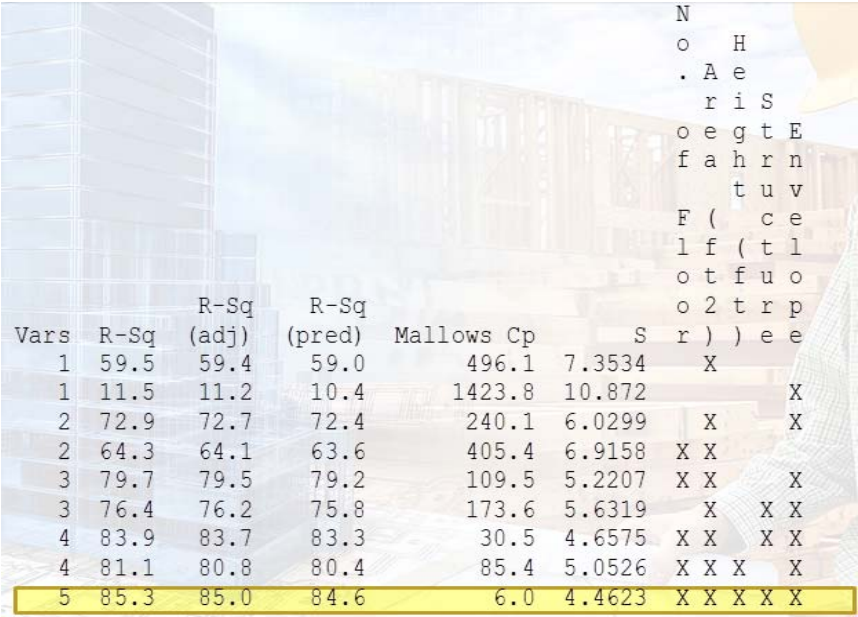
A: Building area in square feet

H: floor height which ranges between 10 to 13 feet

S: Structure type; (1 for wood joist and 2 for steel joist)

E: Envelope type; (1 for Cavity, 2 for Stucco, and 3 for Veneer brick)





Vars	R-Sq	R-Sq (adj)	R-Sq (pred)	Mallows Cp	S	N o H . A e r i S o e g t E f a h r n t u v F (c e l f (t l o t f u o o 2 t r p r)) e e			
1	59.5	59.4	59.0	496.1	7.3534	X			
1	11.5	11.2	10.4	1423.8	10.872				X
2	72.9	72.7	72.4	240.1	6.0299	X			X
2	64.3	64.1	63.6	405.4	6.9158	X	X		
3	79.7	79.5	79.2	109.5	5.2207	X	X		X
3	76.4	76.2	75.8	173.6	5.6319	X		X	X
4	83.9	83.7	83.3	30.5	4.6575	X	X		X
4	81.1	80.8	80.4	85.4	5.0526	X	X	X	X
5	85.3	85.0	84.6	6.0	4.4623	X	X	X	X

Figure 2: Results of best subset analysis based on Minitab.

4.2 Model adequacy assessment tests

Some of the parameters such as response variable (construction costs), floor height, number of floors and structure type have productive correlation among themselves as presented by the developed regression model. On the other hand, a relationship is found to be negative among area of building and envelope type with estimated construction costs.

4.2.1 R^2 and R^2 (adjusted) tests

R^2 and R^2 (adjusted) values are 85.3% and 85.0% respectively as presented by the preliminary tests. The value of R^2 points out that predictor variables describe the variance of 85.3% in response variable (construction costs) for the developed model. The value of R^2 (adjusted) is a changed value of R^2 which conforms too many descriptive terms in the model. Data value for standard deviation (S) is 4.46, and the values of R^2 demonstrate that the built model is much suitable to fit the data.

4.2.2 T-tests

This test is carried out to examine whether all of the predictors are critically correlated with response variables. As shown in Table 1, p-values are 0.000 for estimated coefficient for each predictor. Consequently, alternative hypothesis is accepted while the null hypothesis is rejected. This behavior demonstrates that the response variables “construction costs” are critically correlated with predictors at 0.1 α -level.

4.2.3 F-test

As shown in Table 2, p-value is 0.000 in variance analysis with null hypothesis being rejected. This illustrates the significance of estimated model at 0.05 α -level. As a result, minimum one coefficient is not zero in the developed regression model.

Table 1: Developed model's Statistical diagnostics.

Predictor	Coefficient	SE coef.	T	P	VIF
Constant	166.68	2.86	58.21	0.000	
No. of floor	3.948	0.325	12.15	0.000	1.02
Area (ft ²)	-0.000753	0.000021	-35.17	0.000	1.01
Height (ft.)	1.203	0.234	5.15	0.000	1.00
Structure	5.152	0.571	9.02	0.000	3.60
Envelope	-8.066	0.480	-16.80	0.000	3.63

Table 2: Variance Analysis of developed model.

Source	DF	SS	MS	F	P
Regression	5	32824	6564.8	329.70	0.000
No. of floor	1	2939.5	2939.5	147.62	0.000
Area (ft ²)	1	24626.0	24626.0	1236.75	0.000
Height (ft.)	1	527.3	527.3	26.48	0.000
Structure	1	1620.8	1620.8	81.40	0.000
Envelope	1	5620.4	5620.4	282.26	0.000

4.3 Residual analysis (normality of errors)

The graph for normal probability, demonstrates that all the terms for error are almost distributed normally. Small fluctuations from normalcy are noticed as shown in the graphs (Fig. 3) of residuals' normal probability. These fluctuations are analyzed to be possible unexpected anomalies. By removing these anomalies, the values of certain statistical parameters and R^2 can be improved but the model would remain distant for the perfect demonstration of real world data in hand. As small fluctuations from normality do not demonstrate any serious concerns, the residual analysis results are acceptable [26].

5 VALIDATION OF MODEL

5.1 Plot validation

The validation plot is shown in Fig. 4 in which predicted output is compared with real observations. This illustrates that the predicted values are dispersed nearby real values for response variables. Therefore, the results obtained from first validation test are believed to be acceptable.



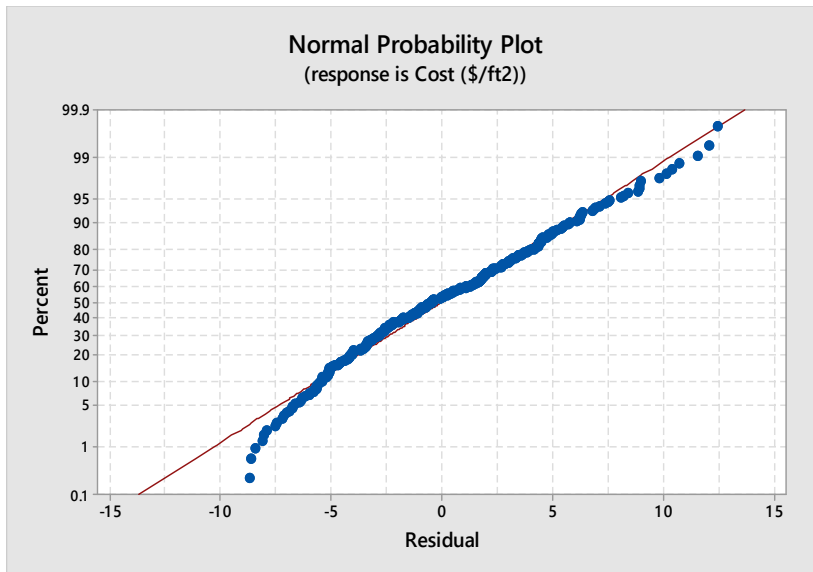


Figure 3: Normal probability plot from residual analysis.

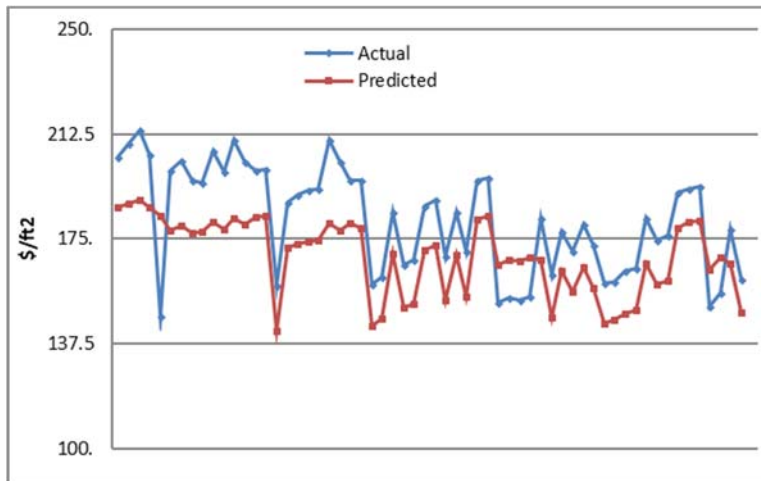


Figure 4: Plot validation for the developed regression model.

5.2 Mathematical validation

$$AIP = \frac{\sum_{i=1}^n \left| 1 - \left(\frac{E_i}{C_i} \right) \right|}{n},$$

$$AIP = \frac{5.6071}{60} = 0.0934$$

$$AVP = 1 - AIP = AVP = 0.9066. \quad (4)$$

The value of AVP depicts the accuracy of predicted model as 90.66%, which is fairly acceptable and satisfactory.

6 CONCLUSION

A regression model has been developed to estimate the construction cost of low-rise residential buildings of maximum 3 floors, suitable to be applied for Canada. A real construction cost data of completed projects was utilized for the evaluation of the national average of construction costs. The input parameters considered were envelope and structure types, number of floors, height of floor and area of building. The model was validated by comparing the predictions with real data of completed projects, as well as by using mathematical formula. The model validation demonstrated acceptable discrepancies while the accuracy obtained by mathematical validation is 90.66%. This model enables buildings owners to assess the economic viability of various structure and envelope types for residential buildings and compare the costs of wood and steel structures. In addition, the model can be used by the owner for budget allocation and cash flow projection for optimum use of resources at the conceptual stage where there is no enough information.

REFERENCES

- [1] Bala, K., Bustani, S. & Waziri, B., A computer-based cost prediction model for institutional building projects in Nigeria: An artificial neural network approach. *J. Eng. Desig and Tech.*, **12**(4), pp. 519–530, 2014.
- [2] Sonmez, R., Review of conceptual cost modeling techniques. *AACE Int. Tran.*, ES71, 2005.
- [3] Sonmez, R., Conceptual cost estimation of building projects with regression analysis and neural networks. *Can. J. Civ. Eng.*, **31**(4), pp. 677–683, 2004.
- [4] Li, H., Shen, Q.P. & Love, P., Cost modelling of office buildings in Hong Kong: an exploratory study. *Facilities*, **23**(9–10), pp. 438–452, 2005.
- [5] Trost, S.M. & Oberlender, G.D., Predicting accuracy of early cost estimates using factor analysis and multivariate regression. *J. Cons. Eng. Manag.*, **129**(2), pp. 198–204, 2003.
- [6] Lowe, D.J., Emsley, M.W. & Harding, A., Predicting construction cost using multiple regression techniques. *J. Cons. Eng. manag.*, **132**(7), pp. 750–758, 2006.
- [7] Siqueira, I., Neural Network-Based Cost Estimating. Master thesis, Concordia University, 1999.
- [8] Günaydın, H.M. & Doğan, S.Z., A neural network approach for early cost estimation of structural systems of buildings. *Int. J. Proj. Manag.*, **22**(7), pp. 595–602, 2004.
- [9] Kim, G., Yoon, J., An, S., Cho, M. & Kang, K., Neural network model incorporating a genetic algorithm in estimating construction costs. *Build & Envi.*, **39**(11), pp. 1333–1340, 2004.
- [10] Yu, W., Lai, C. & Lee, W.A., WICE approach to real-time construction cost estimation. *Auto in Const.*, **15**(1), pp. 12–19, 2006.
- [11] Cheng, M.-Y., Tsai, H. & Hsieh, W., Web-based conceptual cost estimates for construction projects using Evolutionary Fuzzy Neural Inference Model. *Auto in Const.*, **18**(2), p. 164, 2009.



- [12] Cheng, M., Tsai, H. & Sudjono, E., Conceptual cost estimates using evolutionary fuzzy hybrid neural network for projects in construction industry. *Expert Sys. App.*, **37**(6), pp. 4224–4231, 2010.
- [13] Juszczyk, M., The use of artificial neural networks for residential buildings conceptual cost estimation. *Proceedings of AIP Conference*, eds T. Simos, G. Psihoyios & C. Tsitouras, **1558**(1), AIP, 2013.
- [14] Aibinu, A., Dassanayake, D., Chan, T. & Thangaraj, R., Cost estimation for electric light and power elements during building design: A neural network approach. *Eng. Const. Archit. Manag.*, **22**(2), pp. 190–213, 2015.
- [15] An, S.-H., Kim, G. & Kang, K., A case-based reasoning cost estimating model using experience by analytic hierarchy process. *Build. Env.*, **42**(7), pp. 2573–2579, 2007.
- [16] Koo, S., Hong, T., Hyun, C. & Park, S.J., A study on the development of a cost model based on the owner's decision making at the early stages of a construction project. *Int. J. Strat. Prop. Manag.*, **2**, pp. 121–137, 2010.
- [17] Hong, T., Hyun, C. & Moon, H., CBR-based cost prediction model-II of the design phase for multi-family housing projects. *Exp. Sys. App.*, **38**(3), pp. 2797–2808, 2011.
- [18] Ji, S., Park, M. & Lee, H., Cost estimation model for building projects using case-based reasoning. *Can. J. Civ. Eng.*, **38**(5), pp. 570–581, 2011.
- [19] Koo, C., Hong, T. & Hyun, C., The development of a construction cost prediction model with improved prediction capacity using the advanced CBR approach. *Exp. Sys. App.*, **38**(7), pp. 8597–8606, 2011.
- [20] Moon, S.W., Kim, J.S. & Kwon, K.N., Effectiveness of OLAP-based cost data management in construction cost estimate. *Aut. Cons.*, **16**(3), pp. 336–344, 2007.
- [21] Yu, W., PIREM: a new model for conceptual cost estimation. *Const. Manag. Econ.*, **24**(3), pp. 259–270, 2006.
- [22] Sonmez, R., Parametric range estimating of building costs using regression models and bootstrap. *J. Cons. Eng. Manag.*, **134**(12), pp. 1011–1016, 2008.
- [23] Alshamrani, O., Evaluation of school buildings using LCC & sustainability measures – selection of structure & envelope types for schools – sustainable school buildings vs. conventional school buildings, LAP LAMBERT Academic Publishing, 2014. ISBN: 978-33659-59382-6.
- [24] Alshamrani, O., *Construction Cost Prediction Model for Conventional and Sustainable College*.
- [25] Zayed, T. & Halpin, D. Deterministic models for assessing productivity and cost of bored piles. *Con. Manag. Econ.*, **23**(5), pp. 531–543, 2005.
- [26] Kutner, M., Nachtsheim, C. & Neter, J., *Applied Linear Regression Models*, McGraw-Hill/Irwin, 2004.