NRC Research Press

ARTICLE

Cost assessment of construction projects through neural networks

Umit Gulcicek, Omer Ozkan, Murat Gunduz, and Ismail Hakki Demir

Abstract: Construction performance parameters have experienced greater international attention and discussion in recent years. In this study, the change in the load-bearing system cost of a reinforced concrete housing estate building was investigated in relation to the building importance factor, earthquake region, soil type, floor area, and the number of stories. Three different housing estate projects with seven and fifteen stories were investigated. The structural design calculations were performed according to four different soil types, four different earthquake regions, and four different importance factors. With regards to each combination, project costs were calculated. The changes in the cost were examined with artificial neural network and multiple regression models. According to the results of the analysis, the building importance factor was the most significant component, while earthquake region was the second most significant, followed by the factors of the number of floors and the soil type.

Key words: cost estimation, artificial neural network, earthquake region, building importance factor, soil type, housing estate project.

Résumé : Les critères de performance en construction ont fait l'objet d'une forte attention internationale et ils ont été des sujets de discussions au cours des dernières années. Dans la présente étude, le changement dans le coût du système porteur d'un ensemble résidentiel en béton armé est examiné par rapport au coefficient de risque de l'immeuble, à la région sismique, au type de sol, à la superficie de plancher et au nombre d'étages. Trois différents ensembles résidentiels de sept et de quinze étages ont été examinés. Les calculs de conception structurale ont été réalisés pour quatre types de sols différents, quatre régions sismiques et quatre coefficients de risque. Les coûts de projet ont été calculés pour chaque combinaison de facteurs. Les changements dans les coûts ont été examinés à l'aide de réseaux neuronaux artificiels et de modèles de régression multiple. Selon les résultats de l'analyse, le coefficient de risque de l'immeuble était la composante la plus importante, la seconde était la région sismique, suivie du nombre d'étages et du type de sol. [Traduit par la Rédaction]

Mots-clés : estimation des coûts, réseau neuronal artificiel, région sismique, coefficient de risque de l'immeuble, type de sol, ensemble residential.

1. Introduction

Engineering is the art of developing products useful to human beings by employing the principles of science and mathematics as well as experience, judgment, and common practice. The project process can usually be divided into five important phases, i.e., project conception, project design, project construction, occupation, and disposal phase. Project conception is the recognition of a need that can be satisfied by a physical structure. The project design phase translates the primary concept into an expression of a spatial form that will satisfy the owner's requirements in an optimum and economic manner (Abduh et al. 2012). During each of the first three phases, the cost forecasting should be updated depending on the availability of design drawings and specifications. It is during the project conception and project design phases that the feasibility studies and parametric cost estimates are most likely carried out, in which it is a common practice for the owners to acquire estimates from the architects and cost engineers concerning the cost of constructing a proposed project. Normally at this stage, the cost estimate depends on the particular estimator's experience, and imagination, in addition to a wide range of assumptions, including appraisals of previous projects similar in scope (Chen et al. 2011).

Cost forecasting is a fundamental component of many engineering and business decisions, and generally involves the forecasting of production factors like quantity of labor, materials, utilities, floor space, sales, overhead, time and other costs for a predetermined series of time periods (Šiškina et al. 2009). The accuracy of estimation of construction costs pertaining to a construction project is a critical factor in the success of the project. The cost estimation models, which in the early stages estimate the construction costs in light of minimal project information, are useful in the project conception and project design phases of a construction project. Moreover, improved cost estimation techniques introduce a more efficient control system in construction projects.

Accurate and coherent cost estimation plays a significant role in budget projections, contract negotiations, and financial decision making (Apanavičienė and Juodis 2003). The accuracy of an estimate is measured by how well the initially estimated cost compares to the realized total cost of investment (Chen and Chen 2012).

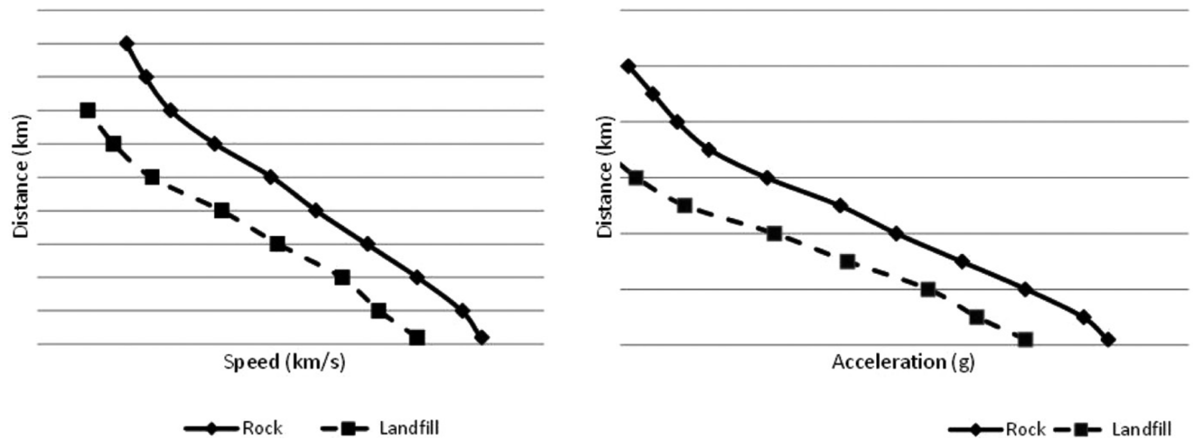
Most building codes over the world require that very important structures be designed in concordance with a seismic coefficient equal to that used for ordinary structures multiplied by a factor greater than one, namely the importance factor. This factor is set intuitively or arbitrarily, varying within a very wide range and it is always independent of the design coefficients of ordinary structures; i.e., the particular site's seismicity and the properties of the

Received 9 October 2012. Accepted 14 April 2013.

U. Gulcicek. Sutcu Imam University, Elbistan Vocational School, 46300 Kahramanmaraş, Turkey.
O. Ozkan. Istanbul Medeniyet University, Faculty of Engineering and Architecture, Department of Civil Engineering, 34720 Istanbul, Turkey.
M. Gunduz. Middle East Technical University, Department of Civil Engineering, 06800 Ankara, Turkey.
I.H. Demir. Sakarya University, Engineering Faculty, Civil Engineering Department, 54187 Sakarya Turkey.
Corresponding author: Murat Gunduz (e-mail: gunduzm@metu.edu.tr).

Published at www.nrcresearchpress.com/cjce on 22 April 2013.

Fig. 1. Effect of distance and soil type.



structures. Thus, the Turkish Building Code (MPS 2007) stipulates factors of 1 or 1.5 depending on the type of facility. A higher factor is used for structures whose failure or collapse might cause a large loss of lives, an extraordinary economic loss, or a loss of irreplaceable structures, and for public buildings that are essential during emergencies (Shen et al. 2007).

First aspect on structural design is to prevent loss of life during earthquakes. For small earthquakes the building should suffer no or minimal damage, for medium earthquakes the building load-bearing system must not get any damage, and for intense earthquakes the load-bearing system must survive. It is a known fact that the earthquake load would affect the building load-bearing system dimensions. The soil type would also affect the expected earthquake load and affect the system dimensions indirectly (Martin et al. 2008). The effect of soil type with respect to distance from epicenter can be seen in Fig. 1. The figure also shows the variation in the velocity and the acceleration values for the same earthquake load.

Numerous factors including but not limited to the structure site's earthquake region (EQ), soil type (ST) and soil parameters, as well as the type of load-bearing material used in the structure and the quality of such materials contribute to the construction costs during the building stage.

In this study, for three different housing projects, the effect of change in earthquake zones and soil types, importance factor, the number of stories and the floor area on cost of building load-bearing system was statistically analyzed. The cost of systems was determined by concrete, steel, and formwork unit prices. The quantity take-off for concrete, steel and formwork were carried out and then transformed into cost by multiplying with the unit prices published by the Ministry of Public Works. The structural design and unit costs were implemented by taking into account the Turkish practice and codes TS498, TS9967, and TS500.

2. Methodology and estimation models

In recent years, various models were developed by many researchers to improve the accuracy and coherence of cost estimations as they greatly influence the budget projections, contract negotiations, and financial decision making. The approaches to cost estimation based on statistics and linear regression analysis have been developed since the 1980s (Tang et al. 2012; Trivailo et al. 2012). Regression, or multiple regression analysis, is a very powerful statistical tool that can be used both as an analytical and predictive technique in assessing the contribution of potential new items to the overall estimate's explanatory power (Li et al. 2009). However, regression analysis is not suitable when describing nonlinear relationships, which are multidimensional and consist of multiple inputs and outputs (Lee 2009).

There is also interest towards applying novel computational techniques, such as fuzzy logic and artificial neural networks (ANN), to the field of cost estimation. Artificial neural networks are purely data driven models, which transition from a random state to a final model via iterative training. Moreover, they do not depend on assumptions about the functional form, probability distribution or smoothness, and are proven universal approximations (Apanavičienė and Juodis 2003).

The structural analysis for three different project types was analyzed by the help of commercial software IDECAD. This software takes into account the necessary parameters required by the Turkish codes TS498, TS9967, and TS500. All projects have seven and fifteen stories and different dimensions with respect to structural members and room sizes. In total, 384 projects were analyzed.

The parameters required for structural design are listed in Table 1. In this table, K_0 is the subgrade reaction, ASP is the allowable soil pressure, T_A and T_B are the spectrum characteristic periods, EQ is earthquake region, I_{factor} is the building importance factor, FA is the floor area, and the FN is the floor number. Structural, dimensional and distributive elements are all responsible for great differences in cost for each zone, for each impact factor, and for each soil type.

After the structural design is carried out by the software, the quantity take-off for concrete, steel and forms was performed. The unit prices for each cost item are then obtained from Turkish Ministry of Housing and Urban Development. The multiple regression analysis and ANN were then carried out for random variables, soil type, earthquake region, importance factor, floor area, and the number of floors.

2.1. Multiple regression model

In statistics, regression analysis includes a number of techniques for modeling and analyzing several variables. The focus may be on the relationship between a single dependent variable and one or more independent variables. More specifically, regression analysis helps one understand how the typical value of the dependent variable changes when any one of the independent variables change.

The general purpose of multiple regression is to gain insight into the relationship between several independent variables and a dependent variable. Multiple regression is a flexible method of data analysis that may be appropriate whenever a quantitative variable (the dependent or criterion variable) is to be examined with respect to its relationship to any other factors (expressed as independent or predictor variables). The relationships may be nonlinear, or the independent variables may be quantitative or qualitative. The effects of a single variable or multiple variables

Table 1. Project parameters.

Soil type (ST)	K_0 (t/m³)	ASP (t/m²)	T_A	T_B	Earthquake region (EQ)	I_{factor}	Square/story (m²) (FA)	Floor number (FN)
ST1	4000	40	0.1	0.3	0.4–0.3–0.2–0.1	1–1.2–1.4–1.5	576.38–358–323.68	7–15
ST2	3000	30	0.15	0.4	0.4–0.3–0.2–0.1	1–1.2–1.4–1.5	576.38–358–323.68	7–15
ST3	2000	20	0.15	0.6	0.4–0.3–0.2–0.1	1–1.2–1.4–1.5	576.38–358–323.68	7–15
ST4	1000	10	0.2	0.9	0.4–0.3–0.2–0.1	1–1.2–1.4–1.5	576.38–358–323.68	7–15

can be evaluated with or without the interfering effects of other variables (Maity 2012).

The relation between building costs and building importance factor, earthquake regions, soil types, construction area, and number of stories were statistically analyzed by commercial software. The dependent variable, which was the building importance factor, earthquake regions, soil types, construction area, and number of stories on a multiple regression analysis. In regression analysis, the value of a dependent response variable is predicted based on the value of independent variables. The model is

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots b_kX_k + e \tag{1}$$

where Y is the independent variable (building cost); b_0, b_1, and b_k are the regression constants; X_1, X_2, and X_k (building importance factor, earthquake regions, soil types, construction area, and number of stories) are the dependent variables; and e is the error. Multiple regression can establish that a set of dependent variables explains a proportion of the variance in an independent variable at a significant level (through a significance test of R^2), and can establish the relative predictive importance of the dependent variables. In our case, building cost was regressed on the dependent variables earthquake zone and soil type. The analysis and the regression function can be seen in Table 2 and eq. (5), respectively.

$$\text{Cost} = 11.125\text{FN} - 12.086\text{EQ} + 5.193\text{ST} + 22.581I_{\text{factor}} + 0.184\text{FA} - 0.016\text{TA} \tag{2}$$

where cost is the cost of load-bearing system, FN is the floor number, EQ is the earthquake zone, ST is the soil type, I_{factor} is the building importance factor, FA is the floor area, and TA is the total area. The variables, constants, and the corresponding t-test are listed in Table 2. If the model fits the data well, the overall R^2 value would be high, and the corresponding p-value would be low. In this case, The R^2 value was 82%, which is high enough to define the building cost. The p-values in Table 2 also indicate that the constant and the predictors are statistically significant.

Analyzing the regression model, it is found that the construction costs are affected by the building importance factor with a coefficient of 22.581, followed by earthquake region (–12.086), number of floors (11.125), soil type (5.193), floor area (0.184), and finally total area (–0.016), from high to low levels of significance factors.

2.2. Artificial neural network

The organization and the functioning of biological neurons inspired the mathematical models that are called the artificial neural networks. The nature of the task that is assigned to the network is related to the presence of numerous artificial neural network variations. Artificial neural networks can automatically approximate whatever functional form best characterizes the data. Artificial neural networks also are inherently nonlinear. That means not only do they estimate nonlinear functions well but also they can extract any residual nonlinear elements from the data after linear terms are removed (Yu et al. 2010). ANN time series model is carried out by means of the following formulation:

Table 2. Multiple regression results.

	FN	EQ	ST	I_{factor}	FA	TA
Coefficient	11.125	–12.086	5.193	22.581	0.184	–0.016
t-value	5.352	–5.431	5.063	10.882	5.853	–2.979
P-value	0.001	0.000	0.001	0.000	0.001	0.000
R^2	0.82		JB		15.25 (0.00)	
Durbin Watson	2.17		LM		1.09 (0.34)	
F_p	0.00		ARCH		0.54 (0.58)	

Table 3. Performance of ANN prediction.

Performance	Per m²
Mean squared error	0.00524
Normalized mean squared error	0.00015
Mean absolute error	0.00523
Min. absolute error	0.00029
Max. absolute error	0.02944
r (linear correlation coefficient)	0.919

$$y_t = G(w_t;\ \psi) = \beta_0 + \sum_{j=1}^{q} \beta_j \psi(\gamma' w_t) + \upsilon_t \tag{3}$$

where $\beta = (\beta_1, \dots \beta_q)'$, $\gamma_j = (\gamma_{j0}, \gamma_{j1}, \dots \gamma_{j,k-1}, c_j)'$, $j = 1 \dots q$; $w_t = (w_{1t}, w_{2t}, \dots w_{kt}, 1)' = (y_{t-1}, y_{t-2}, \dots y_{t-k}, 1)'$, υ_t is neural interpretation diagram (nid), and $\psi(\gamma'_i w_t)$ is the sigmoid function defined as

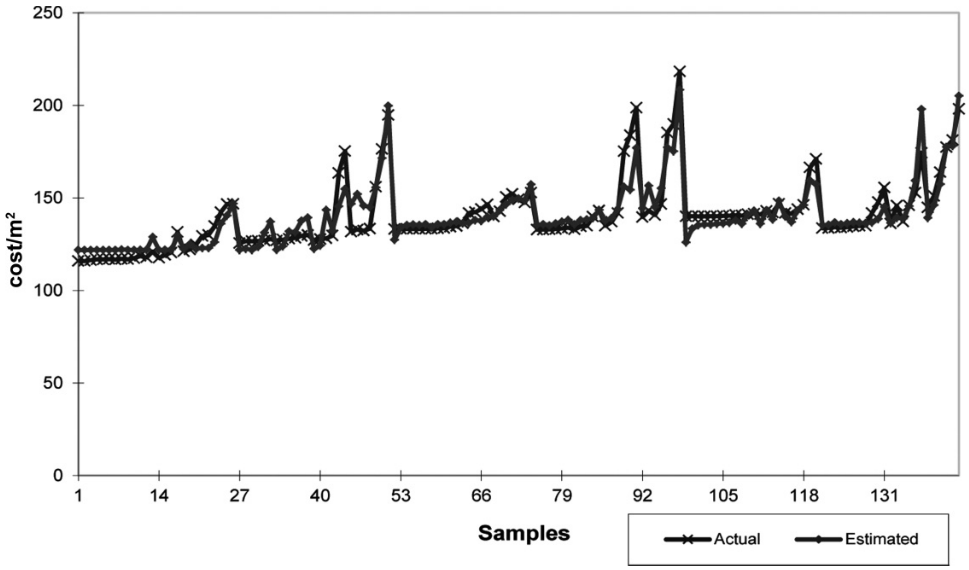
$$\psi(\gamma_j' w_t) = \frac{1}{1 + e^{-\gamma_i' w_t}} \tag{4}$$

In applications to economic time series it is normally useful to include a linear component in eq. (2). The resulting model is defined as

$$y_t = G(w_t;\ \psi) = \alpha_i' w_t + \sum_{j=1}^{q} \beta_j \psi(\gamma_t' w_t) + \upsilon_t \tag{5}$$

where $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k, \alpha_0)0$ is a vector of real coefficients and υ_t is nid. One of the characteristic features of this model is not globally identified. There are three characteristics of neural networks that cause the non-identifiability. The symmetries in the neural network architecture are among the first causes. Permutation of the hidden units will not alter the likelihood function of the model and still there will be q possibilities for every coefficient of the model since $\psi(x) = 1 - \psi(-x)$, a similar outcome can also be observed (Rech et al. 2001). The mutual dependence of the parameters β_i and γ_i, $i = 1, \dots, q$ is the third reason. The γ_i can take any value if corresponding $\beta_i = 0$ and the value of the likelihood function would remain unchanged. Whereas, consistent and asymptotically normal parameters of eq. (4) can be obtained for the maximum likelihood estimates under certain conditions of regularity (Rech 2002).

Fig. 2. The estimated and realized values for the unit cost of construction.



Various researchers have used neural networks as a tool for prediction and optimization previously. But in the area of cost estimating there exist only few applications. The works of Alaei and Salahshoor (2012) and Chen and Huang (2012) in the manufacturing industry, comprise alternative ANN models for cost estimating.

Cheng et al. (2009) developed a system that was based on a neural network that was trained to interpret cost estimates when a new technology was introduced for construction industry. They also found that neural networks outperform linear regression analysis for cost estimation. Cheng et al. (2010) designed a neural network model for early cost estimation of packaging products. In their model, they extracted and quantified cost-sensitive attributes of a product design. The correlation between these cost features and the final cost of the product was found by using a back-propagation neural network algorithm dependent on historical data.

To estimate the effect of structure parameters on the rough construction cost, the ANN model was utilized. Generalized feed forward networks are a generalization of the multiple layer perceptron (MLP) such that the connections can overlap one or more layers. A two-layer feed forward neural network model was developed for prediction of the rough construction cost using an optimized back propagation training algorithm. In this study, the scaled conjugated gradient algorithm was selected as the optimized training method. The data are divided into two parts as the training and the testing sets. Two hundred and forty data are selected as the training set and they are employed to train the ANN based model. One hundred and fourty-four data, which are not used in the training process, are selected as the testing set and they are used to validate the generalization capability of the ANN based model. In the ANN modeling process, the input and the output exchange rate data sets for each parameter are normalized in the following formulation (0, 1). Figure 2 demonstrates the ANN model's training performance with respect to rough construction unit cost.

$$D = 0.1 + 0.8\frac{(D_t - D_{min})}{(D_{max} - D_{min})} \tag{6}$$

where D is the normalized value, D_t is the value to normalize, and D_{min} and D_{max} are the minimum and maximum variable values, respectively.

Figure 3 indicates that the ANN training model of rough construction unit cost reaches the preset training goal of 0.000534 after 1000 iterations.

The statistical parameters of the estimation results obtained from the ANN model in accordance with the learning network created between the economic data and exchange rates are given in Table 3. The estimated and realized values for the unit cost of rough construction are displayed in Table 3.

The verifications stage indicates that the model prediction results reasonably match the observed cost of building load-bearing system. The correlation coefficient between the ANN model predicted values is 0.919. This outcome reveals that the estimation of unit rough construction cost per square metre utilizing the structural parameters was quite powerful.

2.3. Comparison among forecasting models

To evaluate and compare the forecasting performance, it is necessary to introduce a forecasting evaluation criterion. Three different forecast consistency measures are used to compare the performances of the obtained multiple regression and ANN. They are the mean square error (MSE), the mean absolute error (MAE), and the mean absolute percentage error (MAPE). While the above criterions are good measures of the deviations of the predicted values from the actual values, they cannot reflect the ability of a model to predict turning points. The above indicated values are obtained as follows:

$$\text{MSE} = \sum_{t=1}^{T} \frac{(P_t - Z_t)^2}{T} \tag{7}$$

$$\text{MAE} = \sum_{t=1}^{T} \frac{|P_t - Z_t|}{T} \tag{8}$$

$$\text{MAPE} = \frac{1}{T}\sum_{t=1}^{T} \frac{P_t - Z_t}{Z_t} \tag{9}$$

where P_t is the predicted value at time t; Z_t is the actual value at time t; and T is the number of predictions.

The forecasting results of multiple regression and ANN models are summarized in Table 4.

Analyzing the forecasting power of multiple regression and ANNs, it is clearly seen the latter exhibited a much better predic-

Fig. 3. Training performance of the ANN model.

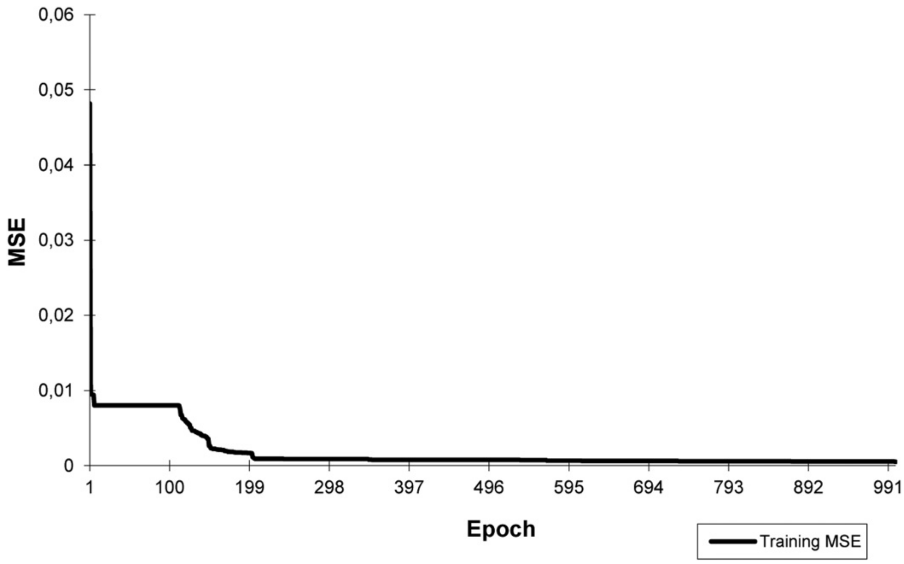


Table 4. Comparison of observed and predicted data from multiple regression and ANN models.

	Multiple regression	ANN
MSE	0.02210	0.00524
MAE	0.01882	0.00015
MAPE	0.01334	0.00523

tion power. It is also emphasized in many studies that ANNs, which are frequently employed in engineering problems, yield a high performance (Apanavičienė; and Juodis 2003; Berlin et al. 2009; Gunduz et al. 2011; Yeh and Deng 2012).

3. Conclusion

In this study, structural parameters contributing to rough estimates of construction costs were analyzed. For this purpose, various parameters such as the number of floors, earthquake zone, soil type, building importance factor, floor area, and total area were investigated. In the first part of the study, multiple regression analysis was performed to evaluate how much these parameters affect the unit cost of construction. According to the results of the analysis, the building importance factor was the most influential component, while earthquake region was the second most influential, followed by the number of floors and soil type. It was found that the floor area and total area have a lower impact on the unit cost compared to the other factors.

In the second part of the study, a prediction model was devised for the structural parameters analyzed as variables. Hence, multiple regression and ANN forecasting models were created, and performance of these models were investigated. According to the analysis, it was inferred that the ANN prediction model displayed a high performance. The approximate unit cost is always a primary concern for construction firms regarding the decision making process in advance of a tender. Inaccurately calculated unit costs adversely affect the profitability and business capabilities of the firms. For instance, overestimated units costs might lead to losing the tender to a competitor, while underestimated unit cots may hamper the probability and the capability to complete the undertaken project. The ANN model used in this study is a significant step towards overcoming the difficulties encountered by the construction firms in the estimation of unit cost. Finally, using the six different variables being analyzed, the prediction of unit cost can be carried out with an accuracy of 90%. The 90% was achieved by comparing the unit cost values to the cost values estimated by the ANN model.

References

Abduh, M., Soemardi, B.W., Cheng, M.Y., Tsai, H.C., and Wirahadikusumah, R.D. 2012. Indonesian construction supply chains cost structure and factors: a case study of two projects. Journal of Civil Engineering and Management, **18**(2): 209–216. doi:10.3846/13923730.2012.671259.

Alaei, H.K., and Salahshoor, K. 2012. The design of new soft sensors based on PCA and a neural network for parameters estimation of a petroleum reservoir. Petroleum Science and Technology, **30**(22): 2294–2305. doi:10.1080/10916466.2010.512899.

Apanavičienė, R., and Juodis, A. 2003. Construction projects management effectiveness modeling with neural networks. Journal of Civil Engineering and Management, **9**(1): 59–67. doi:10.1080/13923730.2003.10531302.

Berlin, S., Raz, T., Glezer, C., and Zviran, M. 2009. Comparison of estimation methods of cost and duration in IT projects. Information and Software Technology, **51**(4): 738–748. doi:10.1016/j.infsof.2008.09.007.

Chen, H.L., Chen, W.T., and Wei, N.C. 2011. Developing a cost-payment coordination model for project cost flow forecasting. Journal of Civil Engineering and Management, **17**(4): 494–509. doi:10.3846/13923730.2011.604540.

Chen, J.H., and Chen, W.S. 2012. Contractor costs of factoring account receivables for a construction project. Journal of Civil Engineering and Management, **18**(2): 227–234. doi:10.3846/13923730.2012.671272.

Chen, T.Y., and Huang, J.H. 2012. An efficient and practical approach to obtain a better optimum solution for structural optimization. Engineering Optimization. In press. doi:10.1080/0305215X.2012.713357.

Cheng, M.Y., Tsai, H.C., and Hsieh, W.S. 2009. Web-based conceptual cost estimates for construction projects using Evolutionary Fuzzy Neural Inference Model. Automation in Construction, **18**(2): 164–172. doi:10.1016/j.autcon.2008.07.001.

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, **37**(6): 4224–4231. doi:10.1016/j.eswa.2009.11.080.

Gunduz, M., Ugur, L.O., and Ozturk, E. 2011. Parametric cost estimation system for light rail transit and metro track works. Expert Systems with Applications, **38**(3): 2873–2877. doi:10.1016/j.eswa.2010.08.080.

Lee, M.C. 2009. Predicting and explaining the adoption of online trading: An empirical study in Taiwan. Decision Support Systems, **47**(2): 133–142. doi:10.1016/j.dss.2009.02.003.

Li, H., Chan, N., Huang, T., Guo, H.L., Lu, W., and Skitmore, M. 2009. Optimizing construction planning schedules by virtual prototyping enabled resource analysis. Automation in Construction, **18**(7): 912–918. doi:10.1016/j.autcon.2009.04.002.

Maity, A. 2012. A powerful test for comparing multiple regression functions. Journal of Nonparametric Statistics, **24**(3): 563–576. doi:10.1080/10485252.2012.677842.

Martin, J.R., Olgun, C.G., Zobel, C.W., and Durgunoğlu, H.T. 2008. Soil improvement for mitigation of damage during the 1999 Kocaeli Earthquake. Journal of Earthquake Engineering, **12**(2): 211–221. doi:10.1080/13632460802014063.

Ministry of Public Works & Settlement (MPS). 2007. Turkish earthquake resistant design code - specification for structures to be built in disaster areas, Ankara, Turkey.
Rech, G. 2002. Forecasting with artificial neural network models. SSE/EFI Working Paper Series in Economics and Finance, No: 491.
Rech, G., Teräsvirta, T., and Tschernig, R. 2001. A simple variable selection technique for nonlinear models. Communications in Statistics: Theory and Methods, **30**(6): 1227–1241. doi:10.1081/STA-100104360.
Shen, L.Y., Hao, J.L., Tam, V.W.Y., and Yao, H. 2007. A checklist for assessing sustainability performance of construction projects. Journal of Civil Engineering and Management, **13**(4): 273–281. doi:10.1080/13923730.2007.9636447.
Šiškina, A., Juodis, A., and Apanavičiene, R. 2009. Evaluation of the competitiveness of construction company overhead costs. Journal of Civil Engineering and Management, **15**(2): 215–224. doi:10.3846/1392-3730.2009.15.215-224.
Tang, S., Wang, D., and Ding, F.Y. 2012. A new process-based cost estimation and pricing model considering the influences of indirect consumption relationships and quality factors. Computers and Industrial Engineering, **63**(4): 985–993. doi:10.1016/j.cie.2012.06.010.
Trivailo, O., Sippel, M., and Şekercioğlu, Y.A. 2012. Review of hardware cost estimation methods, models and tools applied to early phases of space mission planning. Progress in Aerospace Sciences, **53**(1): 1–17. doi:10.1016/j.paerosci.2012.02.001.
Yeh, T.H., and Deng, S. 2012. Application of machine learning methods to cost estimation of product life cycle. International Journal of Computer Integrated Manufacturing, **25**(4–5): 340–352. doi:10.1080/0951192X.2011.645381.
Yu, K.M., Zhou, J., Hong, T.P., and Zhou, J.L. 2010. A load-balanced distributed parallel mining algorithm. Expert Systems with Applications, **37**(3): 2459–2464. doi:10.1016/j.eswa.2009.07.074.