JOURNAL OF CIVIL ENGINEERING AND MANAGEMENT
2009
15(4): 405–409

PREDESIGN COST ESTIMATION OF URBAN RAILWAY PROJECTS WITH PARAMETRIC MODELING

Rifat Sonmez[1] and Bahadir Ontepeli[2]

[1]*Department of Civil Engineering, Middle East Technical University, Ankara, 06531 Turkey,*
E-mail: rsonmez@metu.edu.tr
[2]*Renaissance Construction, 110 Refik Belendir Sok., Ankara, 06540 Turkey*

Received 18 Feb 2009; accepted 07 July 2009

Abstract. This paper presents a parametric modeling method for predesign cost estimation of urban railway systems. Data of 13 light rail and metro projects located in Turkey were compiled for quantification of the impacts of parameters on the project costs. Parametric models have been developed using regression analysis and neural networks techniques. Ten linear regression models were used for determination of the parameters significantly impacting cost of urban railway projects. Two neural networks were considered as an alternative to regression models, particularly for the identification of the non-linear relations. Predictive behaviour and performance of the models were compared to determine a model that presents adequate relations and has a reasonable accuracy. The proposed method provides a powerful approach for determination of a satisfactory parametric cost model during early project stages by incorporating a coordinated use of regression analysis and neural network techniques.

Keywords: predesign cost estimation, urban railway projects, regression analysis, neural networks.

1. Introduction

Cost estimates may be required at early projects stages, before completion of a detailed design for several purposes including budgeting and feasibility decisions. Predesign cost estimates are usually determined by parametric modeling technique during these early project stages, as detailed cost estimation cannot be performed. In parametric modeling, cost of previous similar projects and data of important parameters impacting cost are compiled. The data is used to model the variations in project costs using the parameters. Regression analysis and neural network modeling techniques are generally implemented for quantification of the impacts of parameters on the cost. Key feasibility and budgeting decisions are made based on the parametric cost estimates, although they are usually determined with very limited information.

Parametric models were developed for estimation of predesign construction costs in several studies. There the regression analysis was used for parametric cost estimation of building projects (Kouskoulas and Koehn 1974; Karshenas 1984; Sonmez 2008). Neural networks have been suggested as an alternative to regression analysis for predesign cost estimation of building projects (Gunaydin and Dogan 2004; Kim *et al.* 2005). Parametric models have also been developed for cost estimation of highway projects. Neural network models with 10 parameters were trained by data of 14 projects and, were validated with 4 projects for predesign cost estimation of highway projects (Hegazy and Ayed 1998). The main advantage of neural networks is their capabilities for identifying relations between the parameters and cost. However, the neural network models may have poor prediction performance, especially when complex models are developed with limited data. Therefore, coordinated use of regression and neural network techniques can lead to development of satisfactory parametric models (Sonmez 2004).

The previous parametric modeling studies mainly focused on parametric estimation of construction costs for building and highway projects. Predesign estimation of urban railway system costs is crucial especially in Europe where significant light rail and metro construction is expected. The European Rail Research Advisory Council estimates the potential of new light rail projects as 30 billion EUR, and new metro projects as 95 billion EUR in Europe over the next 15 years (Light … 2004). The growth potential of new urban railway projects is higher especially for emerging countries such as Turkey. The main objective of this paper is development of parametric models for predesign cost estimation of urban railway projects. Regression analysis and neural network techniques were incorporated to achieve a satisfactory parametric model.

2. Project data

Data of 13 urban railway projects were compiled for development of the parametric models (Ontepeli 2005). The projects consisted of 8 metro and 5 light rail transit projects. The length of urban railway systems was between 1.7 to 18.5 km. The projects were located in Turkey and were contracted between 1986–2005. Contract amounts of the projects were escalated using cost index of Minis-

ISSN 1392–3730 print / ISSN 1822–3605 online
DOI: 10.3846/1392-3730.2009.15.405-409

try of Public Works of Turkey. The escalated contract amounts for 2005 were between 50 million TL to 820 million TL (1 EUR = 2.11 TL). The contract amounts did not include the cost of rolling stock and electromechanical systems. All of the projects consisted of double track line railway systems.

Interviews were conducted with the experts from contractors to identify parameters that would impact the cost of urban railway projects. 6 factors were identified as the primary parameters that could impact the projects costs. The first parameter that the experts identified was the percentage of the total length of the tunnel sections executed by TBM boring and/or NATM over the length of the line (PTN). The experts believed that percentage of the tunnel sections had a very significant impact on the unit cost of the urban railway projects. The second primary parameter identified was the percentage of total length of elevated sections over the total length of main line (PES). The percentage of total length of at grade sections over the length of main line (PAG) was the third parameter, and the percentage of total length of tunnel sections executed by cut-and-cover method over the length of main line (PCC) was the fourth parameter. In some projects procurement and installation of the rails were not included in the scope of the contractor. The inclusion of supply and installation of rails (SRW) were identified as the fifth parameter. Finally the sixth parameter was the number of underground stations (UGS) in the project. The experts thought that the number of underground stations could have an impact on the unit cost of the projects.

Five secondary parameters were also identified in addition to six primary parameters. The impacts of secondary parameters on projects costs were not expected to be as significant as the impact of the primary ones. Summary of the secondary parameters are presented in Table 1. Regression analysis was performed to determine the significance of the impacts of primary and secondary parameters on the unit cost of the urban railway projects.

Table 1. Secondary parameters

Parameter	Description
CTP	Contract type (lump sum or unit price)
AGS	Number of at-grade stations
EST	Number of elevated stations
TLM	Total length of main line
PDO	Percentage of total length of depressed-open sections (ramps) to the TLM

3. Regression models

Significance of primary parameters was evaluated in the first part of the regression analysis. All of the six primary parameters were included as independent variables in the first regression model. The cost per km of double track line for year 2005 (UC) was used as the dependent variable in the parametric models. The first regression model (R1) was in the following form:

$$UC = \beta_0 + \beta_1 PTN + \beta_2 PES + \beta_3 PAG + \beta_4 PCC + \beta_5 SRW + \beta_6 UGS. \quad (1)$$

The significance of the parameters in the regression models were evaluated by the *P* values of the regression coefficients corresponding to the parameter. In general, a large *P* value would indicate that the contribution of the parameter to the model might be insignificant. Inclusion of insignificant parameters in the model could lead to poor prediction. Therefore, elimination of insignificant parameters may improve the prediction performance of the models.

The *P* value of the β_4 corresponding to the parameter PCC was 0.92 in the first regression model R1 (Table 2). The value was the largest *P* value in the model, thus, parameter PCC was eliminated. The second regression model (R2) included all of the remaining primary parameters. Results of the second model revealed that the parameter UGS with a *P* value of 0.69 did not contribute to the model significantly, so this parameter was not included in the third regression model. PAG and SRW were the third and fourth parameters eliminated, as the *P* values of coefficients corresponding to variables indicated that the parameters would not have a significant contribution to the models. The fifth regression model R5 included the parameters percentage of tunnels sections (PTN) and percentage of elevated sections (PES). Both parameters contributed significantly as the largest P value in the model was 0.01.

The secondary parameters were not included in the initial regression model R1 as data of only 13 projects were available. The significance of the secondary parameters was determined in the second part of regression analysis; by evaluating one parameter at a time. The first secondary parameter evaluated was CTP. The independent variables of the sixth regression model (R6) consisted of the significant parameters PTN, PES and the parameter CTP.

Table 2. Regression models

Model	Parameter corresponding to the coefficient with the largest *P* value	*P* value of the coefficient
R1	PCC	0.92
R2	UGS	0.69
R3	PAG	0.23
R4	SRW	0.23
R5	PES	0.01
R6	CTP	0.40
R7	AGS	0.33
R8	EST	0.71
R9	TLM	0.66
R10	PDO	0.47

CTP did not contribute to the model, as the P value of the corresponding coefficient was 0.4. The significance of the remaining secondary parameters AGS, EST, TLM and PDO were evaluated in the regression models R7, R8, R9 and R10 respectively. The models consisted of the parameters PTN, PES and the secondary variable being evaluated. None of the secondary variables had a significant contribution as the *P* values of the coefficients were equal or larger than 0.3.

The regression models developed included only linear relations between the parameters and unit cost of urban railway projects. Inclusion of non-linear terms may change the model results significantly. However, when several parameters are included at the same time, decisions related to class of relations between parameters and cost become challenging. Neural networks provide powerful solutions for determining the non-linear relations between the parameters and cost.

4. Neural network models

Neural network models consist of simple computational units organized into a sequence of layers and interlinked by a system of connections. The neural network models have the capability of determining the relations between the input and output parameters. In parametric cost modeling neural networks could quantify the impact of parameters on the cost by capturing the relations existing in the historical project data. A typical parametric neural network model would include the parameters in the input layer, and cost in the output layer. The neural network model should also include at least one hidden layer between the input and output layers to represent the relations between the parameters and cost.

Two feed-forward neural network models were considered as an alternative to linear regression models developed. In the first neural network model (N1) all of the 6 primary parameters were included in the input layer, and escalated unit cost (UC) was included in the output layer (Fig. 1). The neural network N1 had one hidden layer with 7 hidden units (HU). The number of units in the hidden layer was determined by adding the number of input units to the number of output units. The neural network model was trained by back-propagation training algorithm (Rumelhart *et al.* 1986) using data of the 13 projects.

In the second neural network model (N2) only the two parameters that had significant contribution to linear regression models were included. The input parameters consisted of percentage of tunnels sections (PTN) and percentage of elevated sections (PES). The output was the escalated unit cost. One hidden layer with 3 hidden units was included in the second neural network model.

The neural network models contribution in determination of the adequate parametric models is important particularly during determination of potential non-linear relations. However, the neural network models may present certain limitations. One of the main limitations of neural network models may occur especially when complex neural networks are trained with limited amount of data. Neural networks may memorize the data and capture certain complicated relations, which may result in poor prediction performance or inadequate predictive behaviour. This situation could occur in parametric cost modeling, as usually several parameters are considered in the neural network models and limited data are available for training. Comparison of prediction performance of the neural network models with linear regression models would give an indication for the prediction power of the models.

5. Prediction performance of models

Prediction performance of the linear regression model R5 and 2 neural network models were compared to determine a model with reasonable accuracy for parametric estimation of urban railway project costs. A method based on cross validation technique was used for comparison of the prediction performance. First two projects were randomly selected as the first test sample; then remaining projects were used to develop the regression and neural network models. Costs of projects in the test sample were predicted with the models developed. The procedure was repeated for test samples two to six, each containing two different projects selected randomly. There were 3 projects in test sample 6, as the total number of projects was 13. Mean absolute percent error (MAPE) was used to compare prediction performance of the models.

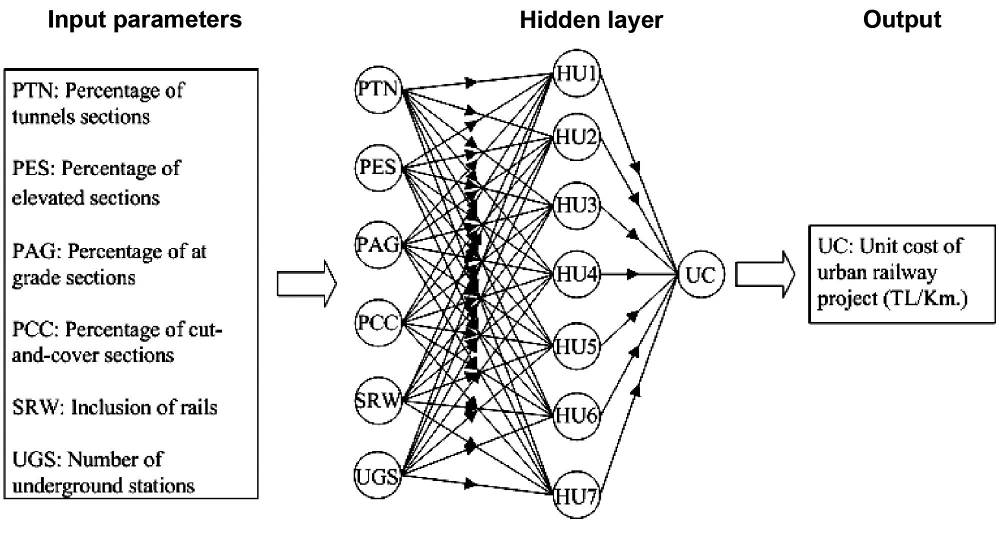

Fig. 1. Neural network model N1

MAPE was calculated as:

$$MAPE = \frac{1}{13}\sum_{i=1}^{13}\frac{|Actual_i - Predicted_i|}{Predicted_i}100. \quad (2)$$

MAPE values of the regression model R5, 6 parameter neural network model (N1), and 2 parameter neural network model (N2) are included in Table 3. N2 had the best prediction performance with a MAPE value of 33.3 among the models. The MAPE value of the regression model R5 was 35.2. The prediction performance of the neural network model with 6 parameters was considerably worse than the performance of N2 and R5.

Table 3. Prediction performance of models

Model	MAPE
R5	35.2
N1	49.8
N2	33.3

The expected accuracy range for feasibility estimates could be considered as –30% to +50% (Cost 1997). The prediction performance of both models N2 and R5 were within the recommended range. The neural network model with 2 significant parameters seemed to have a slightly better performance than the regression model. Neural network and regression models with 2 parameters were determined as the 2 candidate models based on the results of prediction performance comparison. For final model selection predictive behaviour of these two models were analyzed.

6. Predictive behaviour of models

Predictive behaviour of the models could be visualized by sensitivity analysis to reveal how the models capture the impact of parameters on the cost. Sensitivity analysis of the models plays an important role to determine the adequacy of relations suggested by the models. Predictive behaviour of the neural network models is crucial particularly, as the models have the potential to present unnecessary complex relations, especially when they are trained with sparse data.

Sensitivity analysis of models R5 and N2 were performed by varying the parameters PTN and PES. The minimum levels for both of the parameters percentage of tunnels sections (PTN) and percentage of elevated sections (PES) were 0%. The maximum level for PTN was 99%, and was 24% for PES. Outputs of the models R5 and N2 were plotted against varying levels of PTN and PES to compare predictive behaviour of the models. The plot of R5 is given in Fig. 2 and, plot of N2 is given in Fig. 3.

The plot of the R5 illustrates a linear surface as expected, indicating that the regression model suggests a continuous uniform increase in the unit cost as the percentage of tunnels sections and, percentage of elevated sections increase. The neural network model N2 on the other hand presents a non-linear surface. The model suggests that after the percentage of tunnels sections reaches 66%; increasing the percentage of tunnels sections or, percentage of elevated sections would not have an impact on the unit cost. This relation suggested by the neural network model N2 is not adequate, because, unit cost is expected to increase, especially when the percentage of tunnel sections increases. Although the neural

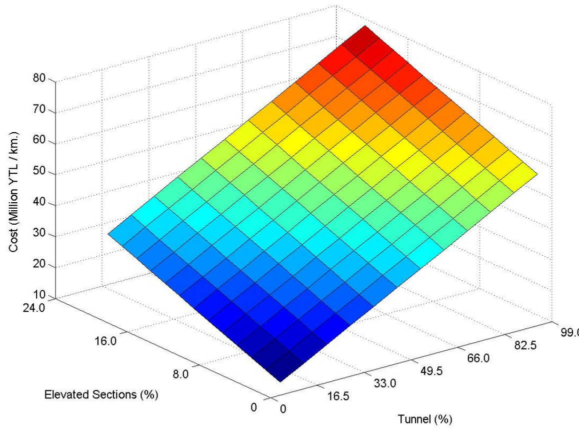

Fig. 2. Sensitivity analysis of model R5

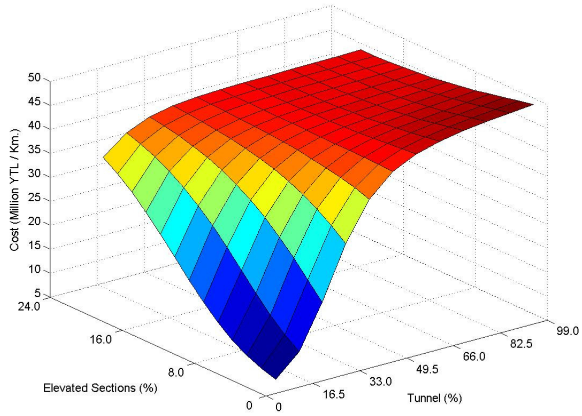

Fig. 3. Sensitivity analysis of model N2

network model N2 has a slightly better prediction performance than the regression model R5, sensitivity analysis reveals that the model is not an adequate one. As conclusion regression model R5 is selected as the satisfactory model for parametric estimation of the urban railway system projects in Turkey.

7. Summary and conclusions

Parametric models were developed for urban railway systems using the data of metro and light rail projects from Turkey. The models could be used to estimate early project costs when construction drawings are not available and detailed cost estimates cannot be performed. Regression analysis and neural network techniques have been implemented for development of parametric models. Depending on the project data, as well as relations between parameters and cost, each technique has certain advantages. Coordinated use of both techniques provides a powerful approach for parametric cost modeling.

Sensitivity analysis reveals crucial information about the predictive behaviour of the parametric neural network models. Although prediction performance of neural network models may be reasonable, the models may suggest inadequate relations between the parameters and cost. Therefore visualization of the predictive behaviour of the neural network models plays an important role for selection of a satisfactory parametric cost model, especially when sparse project data is available.

References

Cost Estimate Classification System Recommended Practice. 1997. No. 187, AACE International, Morgantown, WV.

Light Rail and Metro Systems in Europe. 2004. The European Rail Research Advisory Council, Brussels, Belgium.

Gunaydin, H. M. and Dogan, S. Z. 2004. A neural network approach for early cost estimation of structural systems of buildings, *International Journal of Project Management* 22: 595–602. doi:10.1016/j.ijproman.2004.04.002

Hegazy, T. and Ayed, A. 1998. Neural network model for parametric cost estimation of highway projects, *Journal of Construction Engineering and Management* 124(3): 210–218. doi:10.1061/(ASCE)0733-9364(1998)124:3(210)

Karshenas, S. 1984. Predesign cost estimating method for multistory buildings, *Journal of Construction Engineering and Management* 110(1): 79–86. doi:10.1061/(ASCE)0733-9364(1984)110:1(79)

Kim, G. H.; Seo, D. S. and Kang, K. I. 2005. Hybrid models of neural networks and genetic algorithms for predicting preliminary cost estimates, *Journal of Computing in Civil Engineering* 19(2): 208–211. doi:10.1061/(ASCE)0887-3801(2005)19:2(208)

Kouskoulas, V. and Koehn, E. 1974. Predesign cost-estimation function for buildings, *Journal of the Construction Division* 100(CO4): 589–604.

Ontepeli, M. B. 2005. *Conceptual Cost Estimating of Urban Railway System Projects*. MS thesis, Middle East Technical University, Ankara, Turkey.

Rumelhart, D. E.; Hinton, G. E., and Williams, R. J. 1986. Learning internal representations by error propagation, *Parallel Distributed Processing*, Vol. 1 Foundations (MIT Press, Cambridge, MA).

Sonmez, R. 2004. Conceptual cost estimation of building projects with regression analysis and neural networks, *Canadian Journal of Civil Engineering* 31: 677–683. doi:10.1139/l04-029

Sonmez, R. 2008. Parametric range estimating of construction costs using regression models and bootstrap, *Journal of Construction Engineering and Management* 134(12): 1011–1016. doi:10.1061/(ASCE)0733-9364(2008)134:12(1011)

MIESTŲ GELEŽINKELIŲ PROJEKTŲ PRIEŠPROJEKTINIŲ IŠLAIDŲ SKAIČIAVIMAS, TAIKANT PARAMETRINĮ MODELIAVIMĄ

R. Sonmez, B. Ontepeli

S a n t r a u k a

Pateiktas miesto geležinkelių sistemų priešprojektinių išlaidų skaičiavimo metodas, pagrįstas parametriniu modeliavimu. Parametrų įtaka projekto išlaidoms nustatyta išnagrinėjus 13 nedidelių geležinkelių ir metro projektų Turkijoje. Parametriniai modeliai sudaryti taikant regresinę analizę bei dirbtinio intelekto tinklus. Parametrų įtakos miestų geležinkelių projektavimo išlaidoms reikšmingumui nustatyti sudaryta 10 tiesinės regresijos modelių. Kaip alternatyva regresijos modeliams sudaryti 2 neuroniniai tinklai, ypač nagrinėjant netiesines priklausomybes. Pasiūlytais modeliais gauti rezultatai palyginti tarpusavyje, siekiant nustatyti adekvačias priklausomybes ir užtikrinti reikiamą tikslumą. Pasiūlytas metodas leidžia sukurti parametrinį projekto išlaidų modelį ankstyvojoje priešprojektinėje stadijoje. Tai pasiekta suderintai panaudojus regresinę analizę ir dirbtinio intelekto tinklus.

Reikšminiai žodžiai: priešprojektinių išlaidų skaičiavimas, miesto geležinkelių projektai, regresinė analizė, neuroniniai tinklai.

Rifat SONMEZ. Assoc. Professor, Department of Civil Engineering, Middle East Technical University. Research interests: conceptual cost estimation, construction labour productivity, time-cost-resource optimization.

Bahadir ONTEPELI. Deputy Project Manager, Renaissance Construction. Research interests: conceptual cost estimation.