



HAL
open science

Comparison of linear regression and neural network models to estimate the actual duration of Greek highway projects

Magdalini Titirla, Georgios Aretoulis

► **To cite this version:**

Magdalini Titirla, Georgios Aretoulis. Comparison of linear regression and neural network models to estimate the actual duration of Greek highway projects. XIV Balkan Conference on Operational Research, Virtual BALCOR 2020, Sep 2020, Thessaloniki (virtual), Greece. hal-03178692

HAL Id: hal-03178692

<https://hal.science/hal-03178692v1>

Submitted on 14 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Comparison of linear regression and neural network models to estimate the actual duration of Greek highway projects

Magdalini Titirla

*LMSSC, Conservatoire National des Arts et Métiers, Paris,
France
magdalini.titirla@lecnam.net**

Georgios Aretoulis

*Aristotle University of Thessaloniki, Greece
garet@civil.auth.gr*

Abstract

The aim of this study is to highlight and evaluate parameters that correlate with the actual project duration and compare performance of linear regression with neural network models to predict a reliable final duration for highway construction projects based on data available at the bidding stage. The thirty-seven highway projects that are examined, constructed in Greece, are similar in available data like the extent, the type of work packages and the significance. Considering each project's characteristics and the actual construction duration, correlation analysis is implemented, with the aid of SPSS 25. Correlation analysis identified the most significant project variables. These include archeological findings, type of terrain, land expropriation, the existence of bridge, tunnel and embankment. Furthermore, the WEKA application, through its attribute selection function, highlighted the most efficient subset of variables, both within the initial available attributes but also within the set of attributes identified through correlation analysis. These selected variables through correlation analysis and / or WEKA and appropriate combinations of these are used as input data for linear regression models and neural network models. The corresponding models (linear regression and neural networks) are presented along with their performance. Finally, this paper presents the comparison of linear regression and neural network models to estimate the actual duration of Greek highway projects. Results' discussion and conclusions along with limitations and further research are appropriately analyzed.

KEYWORDS

Highway Construction, Linear Regression, Neural Networks, Predicting Models, Project Actual Duration.

1. INTRODUCTION AND BACKGROUND

Kim et al. (2011), cited in Tsaousoglou et al. (2013) suggest that construction industry plays an important role in leading the national economy and macroeconomic fluctuations substantially influence the construction business. It is therefore essential for each country to promote the implementation of public infrastructure construction projects. During project implementation it is common that variations occur among the initial planning and the actual ongoing project. Variation in project costs usually results in exceeding the initial estimated cost (Aretoulis et al. 2006). At the same time projects suffer from increase in their duration. In this context, it is useful to keep in mind that each highway project is different to another one as there are a lot of parameters which play a role in the differentiation of projects. Research suggests that project location is one of the major differences in construction of highway projects. The unique legal framework in each country influences project implementation, and project procurement systems and contract types (Antonioni et al., 2014). Furthermore, construction projects are influenced by many parameters which alter the cost and schedule planning, both internal and external.

Furthermore, Glymis et al. (2017) based on FANN Tool proposed three, selected, neural network models for the prediction of actual project duration for highways, based on tender budget, length of highway project, number of lanes, number of technical projects, number of bridges, tunnels and road total length. A further research approach focused on highway project duration. The latter is estimated on the basis of variables which were known at the planning phase, and included planned cost, project type and contract type (Irfan et al., 2011). Moreover, Aziz et al. (2007) produced an approach, based on statistical regression analysis. The proposed approach predicts the cost and time of a project and, also, facilitates evaluation of the project's performance during construction. The current study includes, three methods of analysis: ridge regression analysis, general regression analysis and non-linear partial leastsquare regression analysis. More specifically, data collected from completed projects included the type of pavement, contract value, duration and project miles. Jiang and Wu (2007), considered 1818 projects from the area of Indiana. Their research identified factors which affect directly the duration and the cost of each project. These critical factors include type of the project, production rates of the construction firm, location of the project and weather condition.

appeared. Anastasopoulos et al. (2012), used data from 1,722 projects from Indiana. His research identified critical parameters for consideration in models. The variables which were chosen as influential included: project time delay, project cost, project type, weather, level of competition, difference between the winning bid amount and the engineer's estimate and finally difference between the winning bid amount and the next lowest bid.

Pewdum et al. (2009) analyzed highway construction project data. Their research factors affecting project final budget and duration before developing the forecasting models. They employed Artificial Neural Network (ANN). The forecasting results obtained from the proposed method were compared with those obtained from models based on earned value. In a similar context, Hola and Schabowicz (2010) produced a neural network based model to predict productivity for selected sets of machines and to calculate the task execution time and cost. Aretoulis (2019) examined selected public Greek highway projects in order to produce models to predict their actual construction cost based on data available at the bidding stage. Twenty highway projects, constructed in Greece, with similar type of available data were examined. Considering each project's attributes and the actual cost, correlation analysis is implemented, with the aid of SPSS. The optimum neural network model produced a mean squared error with a value of 7.68544E-05 and was based on budgeted cost, lowest awarding bid, technical work cost and electromechanical work cost.

The aim of this study is to highlight and evaluate parameters that correlate with the actual project duration and compare linear regression with neural network models to predict a reliable final duration for highway construction projects based on data available at the bidding stage.

The current research approach also aims at considering special characteristics of projects, realized in Greece. These include archeological findings, type of terrain and land expropriation. Considering each project's characteristic and the actual construction duration, correlation analysis is implemented, with the aid of SPSS 25. This paper presents the results of a relevant literature review on construction project duration, and the recorded corresponding findings, with emphasis on the parameters that have been identified. Furthermore, the WEKA application, through its attribute selection function, highlighted the most efficient subset of variables, both within the initial available attributes but also within the set of attributes identified through correlation analysis. These selected variables through correlation analysis and / or WEKA and appropriate combinations of these are used as input data for linear regression models and neural network models. The corresponding models (linear regression and neural networks) are presented along with their performance. Finally, this paper presents the comparison of linear regression and neural network models to estimate the actual duration of Greek highway projects. Results' discussion and conclusions along with limitations and further research are appropriately analyzed.

2. METHODOLOGICAL APPROACH

The present study is aimed at the production of reliable and efficient models (linear regression and neural networks) for the prediction of actual construction duration of highway projects in Greece. The sample projects under examination include 37 highway projects of similar extent and content (Titirla and Aretoulis, 2019). For these specific projects it became possible to record detailed amount of the same type of data, both quantitative and qualitative. The methodology steps included the following:

Step 1: Consideration of 37 selected highway projects and collection of corresponding data.

Step 2: Construction of an appropriate SPSS database, including both quantitative and qualitative variables.

Step 3: Descriptive statistics of the sample projects' variables.

Step 4: Correlation statistical analysis among the available variables and actual project duration. Analysis was conducted one time for quantitative variables and a second time for mixed variables, both quantitative and qualitative variables.

Step 5: Creation of a ranked list of variables, based on a decreasing degree of correlation coefficients.

Step 6: Proposal of Linear Regression (LR) and Neural Network (NN) models. The models were created starting with the highest correlating variable and then adding one more variable each time from the ordered list based on correlation degree (correlation coefficient) (Step 5).

Step 7: Comparison of the produced LR and NN models.

2.1 Database description

The sample projects under examination include 37 highway projects of similar extent and content. The database included data from 36 Egnatia Odos' sub-projects and 1 smaller project in terms of acreage, cost, and time of completion. Egnatia Odos is one of the most modern highways in the Southeast Europe and part of the TEN-T corridors. The projects were selected on the basis of data uniformity and availability. The available variables are the followings: Project length, Initial cost, Land requirement, Landfill, Tender offer, Embankment, Initial duration, Geotechnical project, Number of lanes, Existence of Tunnel, Number of technical projects, Existence of Bridge. These variables are characterized as quantitative and qualitative and are presented in Table 1. The quantitative variables take on numerical values, and the qualitative variables take on binary (Yes/No) or ordinal values.

Table 1 Variables of investigated projects

Variables	
Quantitative variables	Qualitative variables
Project length	Landfill
Initial duration	Geotechnical project
Initial cost	Embankment
Number of lanes	Tunnel
Number of technical projects	Bridge
Tender offer	Land requirement

2.2 Correlation analysis

An SPSS database was created to record all the available variables. The database consists of 37 cases (number of projects) and 12 variables, which cover common available data across all the projects and according to expert opinion greatly affect the project duration. These variables were recorded as quantitative and also transformed, where possible, into qualitative ones to facilitate additional statistical analyses. Six variables remained as quantitative, while the other six were transformed into qualitative. The values of the IBM SPSS Statistics database were used to identify the correlations among the independent quantitative variables and actual project duration. Furthermore, based on the findings of the correlation analysis, a number of linear regression and neural networks models were designed and implemented, to predict the actual duration of the projects. A brief look on the correlation analyses' results reveals that actual project duration is associated in descending correlation degree with the following variables:

- ✓ initial cost;
- ✓ tender offer;
- ✓ initial duration;
- ✓ embankment;
- ✓ number of lanes;
- ✓ existence of geotechnical projects;
- ✓ length;
- ✓ existence of bridges;
- ✓ and existence of tunnels.

3. MODELS

3.1 Linear regression

Linear regression is a linear approach to modeling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). In linear regression, the relationships are modeled using linear predictor functions whose unknown model parameters are estimated from the data. Such models are called linear models. Most commonly, the conditional mean of the response given the values of the explanatory variables (or predictors) is assumed to be an affine function of those values; less commonly, the conditional median or some other quantile is used. Like all forms of regression

analysis, linear regression focuses on the conditional probability distribution of the response given the values of the predictors, rather than on the joint probability distribution of all of these variables, which is the domain of multivariate analysis. Having identified the potential correlations among the selected common variables, the current paper applies the linear regression analysis. The selected method is the stepwise. The aim is to predict the final duration.

After the analysis of the 37 data cases the following linear regression equations are proposed:

$$\text{Actual project duration} = 841,539 + 4,632 \cdot 10^{-6} \cdot \text{initial cost} \quad (1)$$

$$\text{Actual project duration} = 424,729 + 2,818 \cdot 10^{-6} \cdot \text{initial cost} + 0,843 \cdot \text{initial duration} \quad (2)$$

The first equation includes only one variable, the initial cost and predicts the actual project duration with a predictability of 67% , while the second one includes two variables, the initial cost and the initial duration and it is able to predict the actual project duration with higher predictability equal to 72%. The above-mentioned data are presented in the following tables, namely Table 2 and Table 3.

Table 2 Model Summary from SPSS 25

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	Change Statistics			Sig. F Change	Durbin-Watson
						F Change	df1	df2		
1	.678 ^a	0.460	0.444	296.23948	0.460	29.788	1	35	0.000	
2	.734 ^b	0.539	0.512	277.57713	0.079	5.865	1	34	0.021	1.485

Table 3 Model Coefficients from SPSS 25

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error				Beta	Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance
1	(Constant)	841.539	66.564		12.643	0.000	706.407	976.671					
	INITIAL_COST	4.632E-06	0.000	0.678	5.458	0.000	0.000	0.000	0.678	0.678	0.678	1.000	1.000
2	(Constant)	424.729	183.069		2.320	0.026	52.688	796.769					
	INITIAL_COST	2.818E-06	0.000	0.412	2.579	0.014	0.000	0.000	0.678	0.405	0.300	0.530	1.888
	INITIAL_DURATION	0.843	0.348	0.387	2.422	0.021	0.136	1.550	0.670	0.384	0.282	0.530	1.888

3. 2. Neural networks

Neural networks produced several models. The dependent variable was “Actual Project Duration” in days. Twenty-one projects were used for training the neural network and sixteen projects were used for testing the produced neural network. The methodology involving the application of neural networks for the prediction of actual duration is based on multiple neural networks designed and applied specifically for the problem at hand. The current research is contributing and focusing on identifying the best possible combination of input variables for the optimum prediction. The best Neural network model has the smaller minimum squared error (MSE).

The proposed model is the one that includes the top 12 variables, namely, initial cost, initial duration, length, lanes, technical projects, bridges, tunnels, geotechnical projects, embankment, landfill, land requirement and tender offer. This neural network model has a MSE equal to 1.53E-06.

4. CONCLUSION AND FURTHER RESEARCH

This paper presents the important variables that correlate with the actual project duration and compare linear regression with neural network models to predict a reliable final duration for highway construction projects based on data available at the bidding stage. In total, 37 highway projects constructed in Greece with similar extent and content, were investigated. The corresponding models (linear regression and neural networks) are presented along with their performance. Finally, this paper presents the comparison of linear regression and neural network models to estimate the actual duration of Greek highway projects.

The proposed model of linear regression includes only 2 variables, the initial cost and the initial duration, while for the neural networks the corresponding model includes the top 12 variables, namely, initial cost, initial duration, length, lanes, technical projects, bridges, tunnels, geotechnical projects, embankment, landfill, land requirement and tender offer.

The linear regression model is able to predict the actual project duration with a predictability of 72% while the neural network model produces a MSE equal to 1.53E-06.

A further research in this area could aim at the development and the comparison of the results with other methods and tools. In addition, a wider number of projects should be included, so that correlations and trends that are not observed by this research could be identified in a more extensive research and at the same time increase models' reliability and accuracy.

REFERENCES

Anastasopoulos C.P., Labi S., Bhargava A. and Mannering L.F., 2012. Empirical assessment of the likelihood and duration of highway project time delays. *Journal of Construction Engineering and Management*, Vol. 138, No. 3, pp.390–398.

Antoniou F., Aretoulis G.N., Konstantinidis D. and Papathanasiou J., 2014. Choosing the Most Appropriate Contract Type for Compensating Major Highway Project Contractors. *Journal of Computational Optimization in Economics and Finance (JCOEF)*, Vol. 6, No. 2. pp. 77-95.

Aretoulis G.N., 2019. Neural network models for actual cost prediction in Greek public highway projects. *International Journal of Project Organization Management*, Vol.11, No.1, pp.41 - 64.

Aretoulis G. N., Angelides D. C., Kalfakakou G. P., Fotiadis G. S. and Anastasiadis K. I., 2006. A prototype system for the prediction of final cost in construction projects. *International Journal of Operational Research, Special Issue*, Vol. 6, No. 3, pp. 423–432.

Aziz A.M.A., 2007. Performance Analysis and Forecasting for WSDOT Highway Projects. *Washington State Transportation Commission, Agreement T2695 Task 92*, Washington.

Glymis E., Kanelakis A., Aretoulis G. and Mastoras T., 2017, Predicting highway projects' actual duration using neural networks. In: *LC32017: Volume I – Proceedings of the Joint Conference on Computing in Construction (JC3), July 4-7, 2017*, Heraklion, Greece, pp. 693–699.

Hola B. and Schabowicz K., 2010. Estimation of earthworks execution time cost by means of artificial neural networks. *Automation in Construction*, Vol. 19 No. 5, pp. 570-579.

Irfan M., Khurshid M.B., Anastasopoulos P., Labi S. and Moavenzadeh F., 2011. Planning-stage estimation of highway project duration on the basis of anticipated project cost, project type, and contract type. *International Journal of Project Management*, Vol. 29 No. 1, pp. 78-92.

Jiang Y. and Wu H., 2007. A method for highway agency to estimate highway construction durations and set contract times. *International Journal of Construction Education and Research*, Vol. 3, No. 3, pp. 199-216.

Kim S., Lee S. and Kim J., 2011. Relationship between the financial crisis of Korean construction firms and macroeconomic fluctuations. *Journal of Engineering Construction and Architectural Management*, Vol. 18, No. 4, pp. 407–422.

Pewdum W., Rujiranyong T. and Sooksatra V., 2009. Forecasting final budget and duration of highway construction projects. *Engineering, Construction and Architectural Management*, Vol. 16 No. 6, pp. 544-557.

Titirla M. and Aretoulis G., 2019. Neural network models for actual duration of Greek highway projects. *Journal of Engineering, Design and Technology*, Vol. 17, No. 6, pp. 1323-1339.

Tsaousoglou A. P., Aretoulis G. N. and Kalfakakou G. P., 2013. Forecasting the trends of financial ratios of Greek construction companies. *International Journal of Data Analysis and Information Systems*, Vol. 5, No. 1, pp. 51–63.