International Journal of Advanced Technology and Engineering Exploration, Vol 9(97)
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
http://dx.doi.org/10.19101/IJATEE.2022.10100150

Research Article

Predicting initial duration of project using linear and nonlinear regression models

Gafel Kareem Aswed*, Mohammed Neamah Ahmed and Hussein Ali Mohammed
Assistant Professor, Department of Civil, College of Engineering, University of Kerbala, Karbala, Iraq

Received: 11-September-2022; Revised: 17-December-2022; Accepted: 19-December-2022

Abstract
The fair prediction of completion time of the road project at the initial stage is crucial for the preparation of the tenders. Usually, the engineers using their past experience to assign inaccurate execution time for this type of projects which are not supported by any statistical models. This paper aims to find an acceptable mathematical model that helps specialists in road projects to estimate the implementation time in a more realistic way. Two mathematical models have been developed, which can effectively predict the execution time of road projects in Iraq based on quantities of each of base layer length, width, earthwork, sub base and final cost of the road project. The test results of the nonlinear regression model (NLRM) seem to be more accurate than multi linear regression model (MLRM). NLRM coefficient of determination (R2) is 88.6%, while 76% for MLRM. The mean absolute percentage error (MAPE) 20.29%, 11.65% for MLRM and NLRM respectively. The root means square errors (RMSE) are 76.59, 45.73, and the average accuracy percentages are (AA) 79.71%, 88.35% for MLRM and NLRM respectively. The developed NLRM model reinforces the ability of the decision maker in such projects to accurately estimate the required time pre tendering process.

Keywords
Road duration, Nonlinear model, Project cost, Regression.

1.Introduction
Required time of road implementation is defined as one of the key success factors of the project. The delay issue in the construction industry is a recurring problem that has a negative impact on the project completion [1]. Despite of many advances in the management science, poor performance in the planning phase has been recorded especially regarding the prediction of the road project time. Unrealistic guess in project early phases for the road duration is the primary causes of time overruns [2]. Time overrun in road construction can be led to hamper the prosperity of the national economy when the users of the road must wait for the road delivery more than planned. In spite of challenge, having trustworthy time guess in early phase of project can alleviate these society problems and their impact on the concerned parties. A lot of road projects in Iraq are suffering from not handed over as in planned schedule. It could be said that a project with no delay is not found. From 571 stumbled project in Iraq for the period 2006-2007 it was found that 27% was the road projects [3].

*Author for correspondence

The appropriate guess in the preliminary phase of the project for execution time will minimize the delay in time. Many researchers have been attempted to estimate the required execution time and increase the chance of delivery of the project just in time. Researchers from all over the world, making their best effort to develop mathematical models stepped from simple to more complicated one as explained in the literature review section. This research, participates to the body of knowledge with the proposed road project time estimation methodology over the pre tendering stage that has been lack of interest.

The main objective of this study is predicting the duration of the road projects in Iraq at the pre tendering stage using simple and easy to use models. Two techniques, multi linear regression model (MLRM) and nonlinear regression model (NLRM) have been examined and compared based on the coefficient of determination R2, mean absolute percentage error (MAPE), root means square error (RMSE) and accuracy percentage average (AA) values. Finished road projects in Kerbala province have been used as a case study to achieve the study goal. The clauses of the bill quantities of this type of projects were used as a model inputs. Simulation has been done to create a statistical sample from the same statistical distribution of the original cases.

This paper discussed and review the previous studies in section 2 while section 3 deals with the method of the study, which contains factors that affect the time prediction of the road in the study area, the data simulation. The results of the study and model validations are in section 4 followed by discussions in section 5. Lastly, the conclusions and the suggested future studies presented in section 6 along with study limitations.

2.Literature review

Based on data collected from Indiana Department of Transportation (INDOT) to predict contract time to examine the relation between highway construction project cost and their duration. Factors such as projects region, kind of highways, and weather conditions were used as prediction parameters. They conclude that the contract time could be predicted with 95% confidence interval (CI) using regression analysis. They made adjustments to enclose the impact of the main factors such as road type, location, traffic volume, and execution season on estimated time. The developed model in this study is not for contractor usage, but for the transportation agency [4].

Project type and planned cost which are the available factors at the initial stage of the project were used to estimate the duration from 120 highway project in Pakistan from the year 2001 to 2012. These factors in addition to the project location and their relation to cost overrun risk were used to develop MLRM. Data from four different regions in Pakistan was used to assess the location effect on time risk issue. They concluded that there is a weak correlation between road duration and geographic location of the project. MAPE values ranged from 20% to 40% overestimate or underestimate in project time [5].

Artificial neural network (ANN) model was developed to estimate the duration of prefabricated steel bridge projects on rustic routes in Ghana. Final payment certificate bill of quantity was used and prepared by the ANN. To reduce the factors to a least number of variables, namely the formwork and bridge span as predictors they used a principal component analysis (PCA). They concluded that the in-situ formwork and the bridge span were strongly related to the road duration. Small sample size and no factors other than bill quantity factors are the two limitations of the study [6].

Three MLRM for the highway construction duration in Nigeria were developed. The developed models were in the form of semi-log, linear and log-log transformations using road length, road thickness, no of culverts, and cost/unit length. The results stated that the prediction performance of the log-log model in terms of goodness of fit and prediction accuracy was better than that of the other models [7].

Data from 39 bridge projects in Greece was collected. Using statistical package for social science (SPSS) and Waikato environment for knowledge analysis (WEKA) application to perform correlation analysis between variables. They conclude that the length of the deck, followed by the surface of the bridge and the concrete amount of deck are the most correlated one's with the project duration. Fast artificial neural network tool (FANN) models were created using the sift data from the previous step. Small sample size and model variables were the two limitations of this research. They suggested to adopt another modelling method to compare the results [8].

Depending on six factors which are; road length, number of lanes, number of intersections, earthworks, type of pavement and furniture level ANN model was developed to predict the road duration in Iraq. A significant correlation (90.6%) was found between actual and predict duration of the road project and AA of 74.27%. The results of the study were limited to construct the project by roads and bridges directorate in Iraq between the years 2011 and 2012 [9].

Two variables, the initial cost and the initial duration were used in MLRM to predict the actual duration of the road project using data from 37 projects in Greece at the bidding stage. ANN model was also built using factors such as initial duration, initial cost, length, lanes, embankment, existence of bridges, geotechnical projects, tender offer, technical in projects and landfills. The MLRM models successfully predicted the execution period of the road project with R2=72%, while the ANN model RMSE equal to 1.53E-06. They suggested to include further projects to resolve the correlation issues which are not checked in their study [10].

Based on the available information at the letting stage a hazard-based model conducted to predict the duration of a highway project. The study used survival analysis and (Kaplan–Meier) method to build their models. Project groups are identified when it has a similar survival distribution using the log-rank test. Time-to-event parametric models were developed to connect the project time to the external factors. The models depict the nonlinear pertinence between project time and the external independent variables such as, location of the project, bidding time, source of design and the contractor They concluded that from five models tested that Weibull model is the best one for predicting construction time of bridges projects and traffic work categories while log-logistic is the best for other work categories [11].

Five MLRM relationships were tested to predict the preliminary road construction time in the west bank in Palestine using bid quantities, length and width of road and area as predictors. The MAPE of the constructed models ranged from 19.1% to 31.4%. The study concluded that the inclusion of bid quantities as predicted in the model makes it better than ones include the size of the project. The researcher suggested that the responsible authorities in the construction industry in Palestine must continuously update the productivity data of machines and labour to improve the accuracy of the prediction of the road project duration [12].

The performance of MLRM and ANN models was compared depending on the available factors at the bidding stage and illuminate the more correlated factors with the road duration. Four analytical tools such as time series analysis, ANN and smoothing techniques and Brimelow's model of time–cost (BTC) was used to predict the construction duration of India's government infrastructure projects (highway road projects). The study depended on; the highway status, type, planned duration, time overrun and contract cost as predictors [13].

A study on a rehabilitation road project in Karbala city in Iraq was done. The main activities in the road were; municipal works, sewer works, trench works, optical cable works, water works. Six main activities of the selected road were chosen and it was presented to a group of experts in scheduling such activities. The study assured that the amended project evaluation and review technique (PERT) method using the fuzzy Delphi (FD) process was the most appropriate technique to predict the execution time of the road [14].

MLRM and ANN models were developed using data from Greek highway projects. The data was extracted and tested for correlation analysis. Modelling was starting with variable of heigh correlation value and adding one by one based on their correlation degree. Depending on the value of the correlation coefficient, the variables were entered into the models one after the other starting from the highest correlation. Project length, initial cost, land requirement, landfill, tender offer, embankment, initial duration, geotechnical project, number of lanes, existence of a tunnel, a number of technical projects, existence of the bridge are the study factors. The study revealed that the ANN model is more accurate than MLRM in guessing road project executed time. The small sample size is the main limitation of this study, which is not more than 37 road projects [15].

Data from 100 public road projects finished between 2003 and 2008 in three adjacent cities in southeastern Poland was used for predicting the road duration at the visibility stage. Factors such as cost, length, culverts, bays, intersections, were adopted in this study. They construct three models, simple regression, a multifactor regression and a regression tree and stated that the models were statistically correct but low precise [16].

There is no previous study, to the knowledge of the researchers, that dealt with predicting the duration of the road project in Iraq using simplified models except for one study that used artificial neural networks [9] and the resulted model was complex and difficult to use. In this research some data in the bill of quantity will be used to predict the road duration before the bidding phase. A complete list of abbreviations is shown in *Appendix I*.

3.Methods

3.1Identification of factors

The most recent studies were explored to recap the factors affecting the preliminary guess of road project duration in the planning phase when not enough data is available. The Google scholar search engine was used to check the universal database about the issues. Studies for instance, Mohamed and Moselhi [17], Sharma et al. [18], Le et al. [19], Nevett et al. [20], Son et al. [21], Okere [22], and Nani et al. [23] used multi factors that could be adopted to forecast road duration at the early stage of the project. This study adopted the available factors after preparing the bill of quantity by the estimator and pricing the whole road activities. To achieve the study goal the researcher, collected the data of the completed road projects in order to build the predictive models of road duration. These factors are; length of the road (L) in (m), width (W) in (m), actual duration (D) in days, final total cost

(C) in Iraqi dinars (IQD), barriers cost (B) in IQD such as presence of obstacles such as water lines, sewers and electrical cables, rock (R) in (m^3), earthwork (E) in (m^2), culverts (CL) in m., subbase (S) in m^2 and lastly, asphalt (AS) in m^2 as in *Appendix II*.

3.2Study area

This research is conducted in Karbala province, Iraq. Kerbala is located in the central part of the country, southwest of the capital, Baghdad, at a distance of (110 km). The total area of city is about (5,560 km), which constitutes (1.1%) of the total area of Iraq. More than1500 km length from highway and secondary roads in Karbala Provence. The percentage of paved roads in the governorate is about 85% until 2009, and the unpaved roads represent part of the secondary and rural roads. The population, which was around 1.22 million in 2018 with growth rate 2.1%, it is expected to reach around 3 million by 2025 due to random migration. This situation urged the authorities at the city to open new roads and new districts to overcome the unplanned immigration.

3.3Data simulation and methods

Simulation is a method of statistical sampling used to approximate solutions to quantitative factors. Probability distribution function (PDF) for these factors is determined first. The values of these factors are simulated many thousands of times and each time random numbers are chosen subjected to the same probability distribution for that factor and the resulted data is a new probable project [24]. Collecting fourteen completed projects' data from Directorate of Roads and Bridges of Karbala Governorate in Iraq. Examining the distribution of each variable as a first step to simulate the data. Fifty thousand cases of simulated data have been made using SPSS program as shown in *Figure 1*. Fourteen completed projects' data from Directorate of Roads and Bridges of Karbala Governorate in Iraq. Examining the distribution of each variable as a first step to simulate the data. Fifty thousand cases of simulated data have been made using SPSS program as shown in *Figure 1*.

Linear regression model assumptions were checked using SPSS program. If collinearity between variables are exists it leads to eliminate the corelated ones as in *Figure 2*.

After that NLRM has been tested. Finally, testing the developed models to check its efficiency.

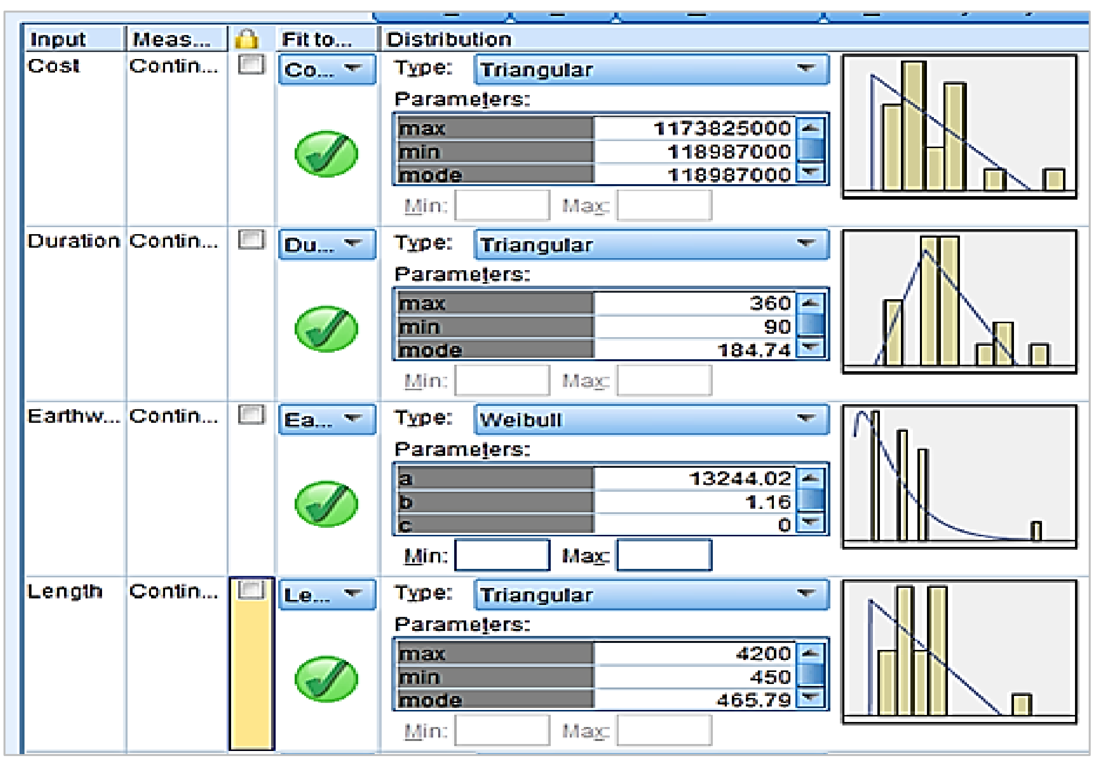

Figure 1 Simulation of the data in SPSS program

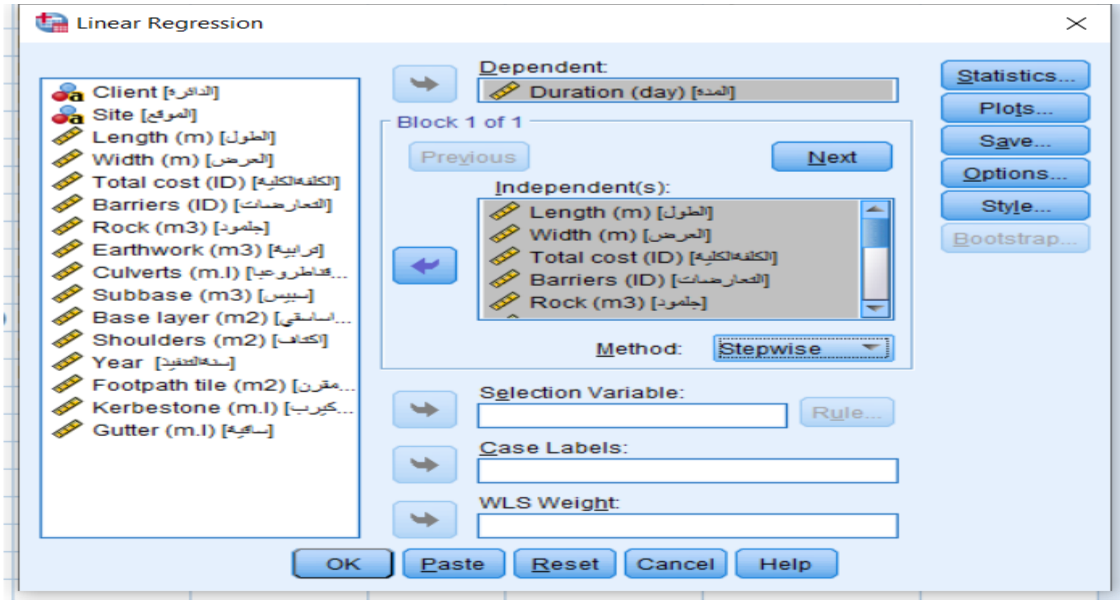

Figure 2 Testing linear regression model assumptions

3.4Non-linear model Test

Nonlinear regression is a system of detecting a nonlinear model of the relation between the dependent variable and a set of independent variables. Even when a linear approach doing well, a nonlinear model could be used to explore a fine performance of the parameters [25]. Nine models (quadratic, s-curve, logarithmic, inverse, power, logistic, growth, compound and exponential) were checked. The favored model between (duration) and each one from the six factors, (Subbase quantity in m^3, road length in m, total cost in IQD, base layer in m^2, earthwork m^3, and road width in m) have been examined. Utilizing SPSS program version 26 by using the previously simulated fifty thousand cases to discover the best fit model based on the maximum R^2 as shown in *Figure 3*. The conceptual framework of the study started from problem identification, literature review, data collection, data simulation, linear and nonlinear models testing, evaluations and discussed as in *Figure 4*.

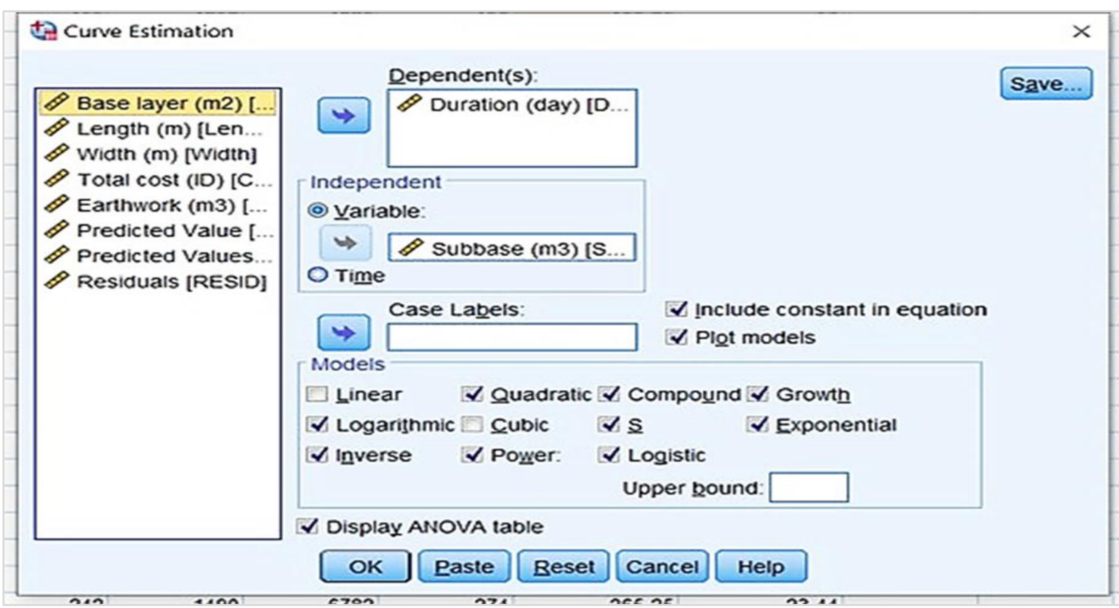

Figure 3 Nonlinear model estimation in SPSS program

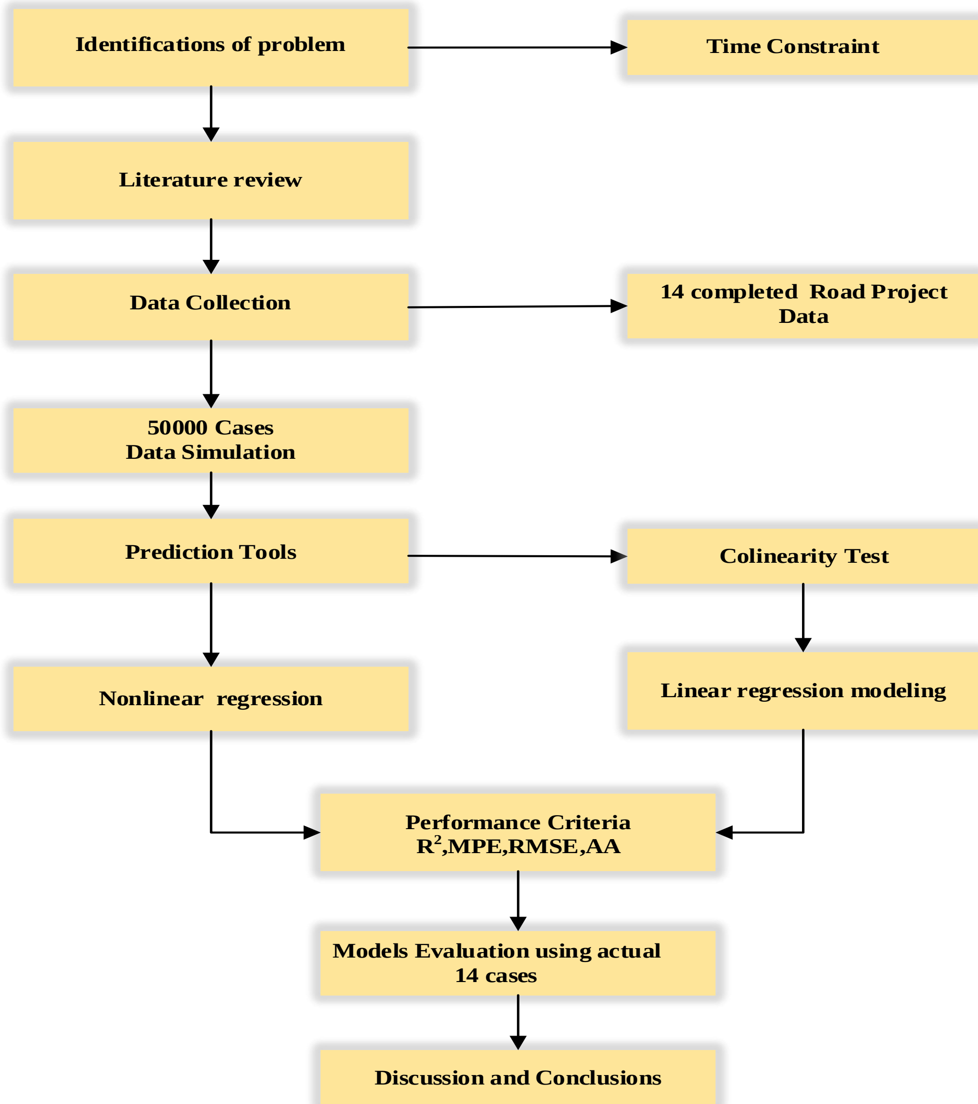

Figure 4 Methodological approach for road duration prediction

4.Results

4.1Linear model development

The MLRM technique is a powerful statistical tool for estimating the functional relationship between a response variable (dependent variable, Y) and one or more explanatory variables (independent variables, X1, X2, X3,….,). It is concerned with finding the best-fit representation for the relationship that exists in the data sample. Regression model must be met the assumptions of linearity, normality and homoscedasticity. MLRM estimation models are well established and widely used in project duration estimation the selected variables as predictors for road total duration are (Subbase quantity in m^3, road length in m, total cost in IQD, base layer in m^2, earthwork m^3, and road width in m). Fifty thousand simulated Cases with 95% confidence interval and forward stepwise method has been chosen to develop the linear analysis. To construct the MLRM, SPSS version 26 is used. Basically, MLRM are intended to find the linear combination variables which best correlate with independent variables [26]. It is found there was A collinearity problem between some inputs, (Subbase, base layer and the total cost) where the tolerance values less than (0.1) and the variance inflation factor (VIF) exceeds (10) typically indicates a multi-collinearity problem [27]. The remaining Three variables (length, width and earthwork) were used to develop the MLRM where there has no collinearity. Model parameters along with the VIF statistics are shown in *Table 1*. Residual tests indicate the normal probability test with mean ≈ zero and one standard deviation. The resulted Equation 1 below with (R^2=82.3%) is:

D=b0+b1×L+b2×w+b3×E (1)

Where; b0, b1, b2, b3 are constants.

The length of road is more important factor in predicting the road duration where Beta value is the greatest and equal to 0.890. All three variables are statically significant based on significance (sig.) values as in *Table 1*.

Table 1 Linear regression noncorrelated predictors

Model	Unstandardized coefficients standardized coefficients			t	Sig.	Collinearity statistics	
	B	Standard Error	Beta			Tolerance	VIF
(Constant)	153.499	2.952	-	51.995	0.000	-	-
Length [m]	0.056	0.000	0.890	409.152	0.000	0.746	1.340
Width [m]	7.776	0.552	0.029	14.083	0.000	0.850	1.177
Earthwork [m3]	0.000	0.000	0.071	35.198	0.000	0.869	1.151

4.2Non-linear model development

To construct the final model, the NLRM analysis in SPSS was used. The best equations of the relation between the duration of the road and each of specific variable resulted from previous analysis in section 3.4 were select and used to build the final NLRM by as in *Figure 5*.

The final mathematical model as in Equation 2.

D=[(b0×B^b1)+(b2+b3×L)+(b4×L^2)+(e^b5+b6/W)+(b7×C^8)+b9+(b10×E)+(b11×E^2)+(b12×S^b13)] (2)

In which D is road duration in days, b0, b1, ……, b13 are the final model constants. Analysis for nonlinear model shows that the R^2=85.8% as shown in *Table 2.* ANOVA analysis for nonlinear model shows the parameters estimate from b0 to b13 along with its standard errors and 95% confidence intervals as in *Table 3.*

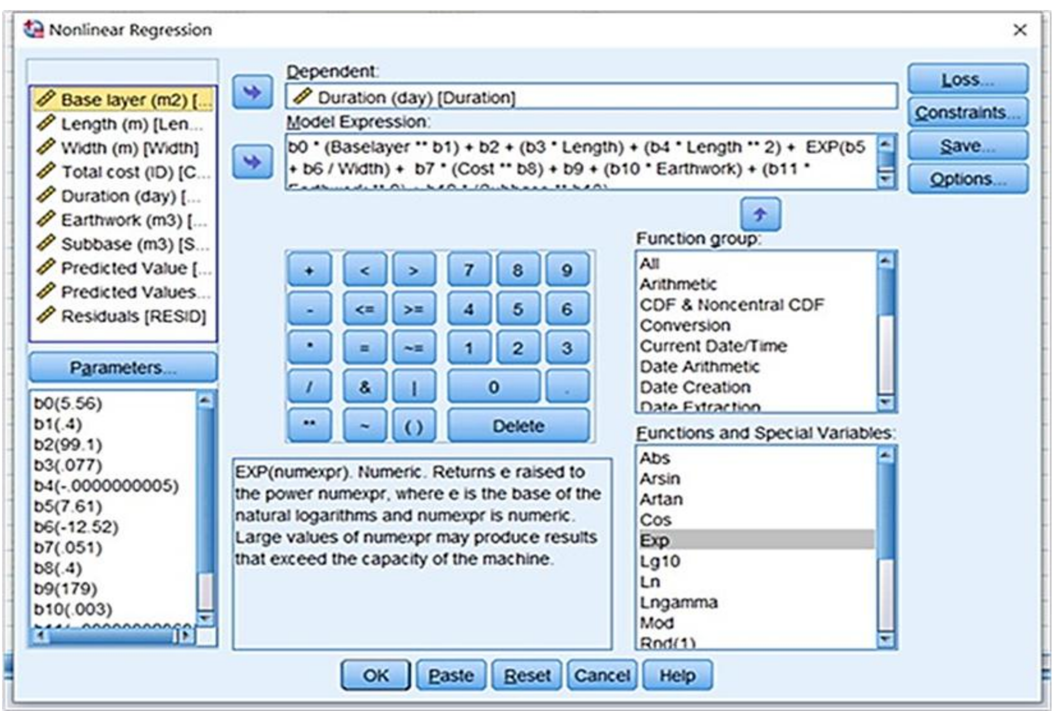

Figure 5 Final nonlinear model in SPSS program

Table 2 Results of analysis of variance for the (NLR) model

Source	Sum of squares	Degree of freedom (df)	Mean squares
Regression	2369488664.411	14	169249190.315
Residual	22145566.235	49986	443.035
Uncorrected Total	2391634230.646	50000	-
Corrected Total	155491157.839	49999	-
Dependent variable: Duration (day)			
R squared = 1 - (Residual Sum of Squares) / (Corrected Sum of Squares) = **0.858**.			

Table 3 Nonlinear regression model coefficients

			95% Confidence interval	
Parameter	**Estimate**	**Standard Error**	**Lower bound**	**Upper bound**
b0	9163.397	642728.014	-1250590.467	1268917.261
b1	0.002	0.127	-0.248	0.251
b2	-4937.278	20503881.760	-40192767.458	40182892.901
b3	0.047	0.001	0.044	0.049
b4	1.722E-6	0.000	1.254E-6	2.190E-6
b5	-6.342	805299.101	-1578401.297	1578388.612
b6	-75.244	5522077.520	-10823406.953	10823256.465
b7	361.608	1541.990	-2660.708	3383.925
b8	0.034	0.088	-0.138	0.206
b9	-4923.430	20519914.577	-40224178.105	40214331.244
b10	0.001	0.000	0.001	0.001
b11	-7.616E-9	0.000	-8.566E-9	-6.667E-9
b12	-0.026	0.008	-0.041	-0.010
b13	0.925	0.032	0.862	0.989

4.3Model validation

Test of the accuracy and validity is the most important step when developing regression models. The developed model is tested and evaluated with some test or validation data. The data used to validate the developed model is from the same population but was not previously used in developing the model. To test the validity of the developed models the original Fourteen actual cases were selected which are not used in the model training as shown in *Appendix II*. The predicted duration values of these projects (computed using the developed mathematical models) were compared to the actual duration and the results of coefficient of determination (R^2=76%, 88.6%) for the MLRM and NLRM respectively as shown in *Table 4*. Using Equation 3 to calculate the MAPE= 20.29%, 11.65%), Equation 4 to calculate RMSE= 76.59, 45.73 and Equation 5 to check the AA= 79.71%, 88.35% for the MLRM and NLRM respectively [28]:

$$MAPE = (\sum^{n} | A-E|/A\times 100)/n. \quad (3)$$

$$RMSE= Sqrt\ [(\sum^{n}(E-A)^2/n] \quad (4)$$

$$AA\%=100\% - MAPE \quad (5)$$

where A is actual value, E is predicted value and n is the total number of cases.

Another check on the linear and nonlinear models' efficiency is the probability plot for the residuals as shown in *Figure 6* and *Figure 7* respectively. The graphs show the goodness of residuals distribute over the two sides of the regression line.

Table 4 Predicted values for project duration

Actual duration (days)	Linear model predicted duration (days)	Non-linear model predicted duration (days)
210	284.19	220.32
180	252.16	231.26
180	213.37	195.42
270	296.13	260.75
120	187.21	163.11
300	288.88	273.06
240	254.77	236.14
180	230.45	213.22

Actual duration (days)	Linear model predicted duration (days)	Non-linear model predicted duration (days)
180	212.96	205.02
240	290.44	261.89
300	288.88	273.06
240	254.77	236.14
180	214.19	189.42
180	212.96	205.02

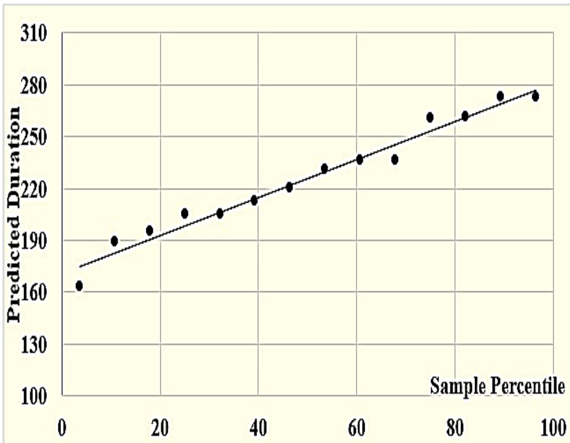

Figure 6 Test of normality for the linear model

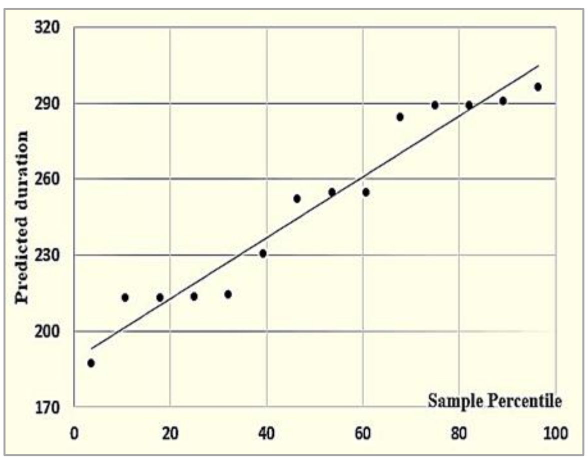

Figure 7 Test of normality for the nonlinear model

5.Discussion

Due to small sample, simulation has been done to generate random cases (projects) according to each of the actual variable statistical distributions. Residual analysis has been performed to evaluate the assumptions of linearity, normality and homoscedasticity in the linear model variable, then excluding factors related to violation of linearity assumptions. The best factors considered for the prediction of road duration at the pre tendering stage using MLRM are Length, Width and Earthwork. The length of the road is the more important factor in the model based on Beta value which is equal to 0. 890. All these variables have a heigh significance value (0.0001) and very small standard errors. The variables used in this model, interpret 76% of the variance in the duration of the road project, therefore the conclusion is that the model showed a moderate agreement when compared with the practical data and heigh AA= 79.71%. The results consent with Mahamid [12] and Waziri et al. [7] studies in Palestine and Nigeria respectively. In the developed NLRM, the considered factors are the quantities of each of base layer, length, width, earthwork, sub base and final cost of the road project. These factors explain 88.6% from the duration variance of the road project and AA=88.35%. The results of the two model's validation indicated that the NLRM is more accurate than MLRM one in terms of R^2, MAPE, RMSE and AA values.

Limitations: This study is limited to the roads in the city of Karbala holy shrines in Iraq that executed between the year 2010 to 2019 by the Directorate of Roads and Bridges. Also, limited to the technique's MLRM and NLRM.

6.Conclusion and future work

The condense of this study was to develop models for estimation of the expected duration of road projects in Iraq. Preparing of reliable time estimates can be helpful to the concerned to deliver optimum project schedules and to avoid the problem related to time overruns. The developed models provided interesting information about the parameters that could be used as predictors of road time before bidding stage. The models are useful to both owners and contractors because of its ability to predict road duration at the earliest project stage and to improve the process of delay alleviation which can ultimately result into more competent road project schedule.

Future Scope: Due to the simplicity of the models, it could be handled using Microsoft excel or any easy software. The same procedures may be adopted to predict the duration of other kind of projects in Iraq. Additional variables may be included like the expected situation of weather based on [3] suggestion. Modeling can be done by other methods as well, such as earned value management, support vector machine, random forest, etc.

Acknowledgment
The authors would like to thank Directorate of Roads and Bridges of Karbala Governorate in Iraq for their cooperation with us and providing executed road project information.

Conflicts of interest
The authors have no conflicts of interest to declare.

Author's contribution statement
Gafel Kareem Aswed: Analysis, Interpretation of results and approved the final version of the manuscript. **Mohammed Neamah Ahmed:** Data collection and preparation. **Hussein Ali Mohammed:** Supervision and Draft manuscript preparation.

References
[1] Khalid FJ. The impact of poor planning and management on the duration of construction projects: a review. Multi-Knowledge Electronic Comprehensive Journal for Education and Science Publications. 2017; 2:161-81.
[2] Khaleel TA, Hadi IZ. Controlling of time-overrun in construction projects in Iraq. Engineering and Technology Journal. 2017; 35(2 Part A):111-7.
[3] Al HBI. An investigation into factors causing delays in highway construction projects in Iraq. In MATEC web of conferences 2018 (pp. 1-11). EDP Sciences.
[4] Jiang Y, Wu H. A method for highway agency to estimate highway construction durations and set contract times. International Journal of Construction Education and Research. 2007; 3(3):199-216.
[5] Kaleem S, Irfan M, Gabriel HF. Estimation of highway project duration at the planning stage and analysis of risk factors leading to time overrun. In T&DI congress: planes, trains, and automobiles 2014 (pp. 612-26).
[6] Mensah I, Nani G, Adjei-kumi T. Development of a model for estimating the duration of bridge construction projects in Ghana. International Journal of Construction Engineering and Management. 2016; 5(2):55-64.
[7] Waziri BS, Jibrin AT, Kadai B. Functional duration models for highway construction projects in Nigeria. Arid Zone Journal of Engineering, Technology and Environment. 2014; 10:13-23.
[8] Karadimos P, Anthopoulos L. Neural network models for actual cost and actual duration estimation in construction projects: findings from Greece. International Journal of Structural and Construction Engineering. 2021; 15(5):250-61.
[9] Al-sadi AM, Zamiem SK, Al-jumaili LA. Estimating the optimum duration of road projects using neural network model. International Journal of Engineering and Technology. 2017; 9(5):3458-69.
[10] Titirla M, Aretoulis G. Neural network models for actual duration of Greek highway projects. Journal of Engineering, Design and Technology. 2019; 17(6):1323-9.
[11] Qiao Y, Labi S, Fricker JD. Hazard-based duration models for predicting actual duration of highway projects using nonparametric and parametric survival analysis. Journal of Management in Engineering. 2019; 35(6):1-16.
[12] Mahamid I. The development of regression models for preliminary prediction of road construction duration. International Journal of Engineering and Information Systems. 2019; 3(4):14-20.
[13] Velumani P, Nampoothiri NV, Urbański M. A comparative study of models for the construction duration prediction in highway road projects of India. Sustainability. 2021; 13(8):1-13.
[14] Nemaa ZK, Aswed GK. Forecasting construction time for road projects and infrastructure using the fuzzy PERT method. In IOP conference series: materials science and engineering 2021 (pp. 1-13). IOP Publishing.
[15] Titirla M, Aretoulis G. Comparison of linear regression and neural network models to estimate the actual duration of Greek highway projects. In XIV Balkan conference on operational research, virtual BALCOR 2020 (pp. 2-6).
[16] Czarnigowska A, Sobotka A. Estimating construction duration for public roads during the preplanning phase. Journal of Engineering, Project, and Production Management. 2014; 4(1):26-35.
[17] Mohamed B, Moselhi O. Conceptual estimation of construction duration and cost of public highway projects. Journal of Information Technology in Construction. 2022; 27(29):595-618.
[18] Sharma S, Ahmed S, Naseem M, Alnumay WS, Singh S, Cho GH. A survey on applications of artificial intelligence for pre-parametric project cost and soil shear-strength estimation in construction and geotechnical engineering. Sensors. 2021; 21(2):1-44.
[19] Le C, Wai YM, David JH, Choi K. Comprehensive evaluation of influential factors on public roadway project contract time. Journal of Management in Engineering. 2021; 37(5).
[20] Nevett G, Goodrum PM, Littlejohn RL. Understanding the effect of bid quantities, project characteristics, and project locations on the duration of road transportation construction projects during early stages. Transportation Research Record. 2021; 2675(2):121-34.
[21] Son J, Khwaja N, Milligan DS. Planning-phase estimation of construction time for a large portfolio of highway projects. Journal of Construction Engineering and Management. 2019; 145(4):1-11.
[22] Okere G. An evaluation of a predictive conceptual method for contract time determination on highway projects based on project types. International Journal of Civil Engineering. 2019; 17(7):1057-73.

[23] Nani G, Mensah I, Adjei-kumi T. Duration estimation model for bridge construction projects in Ghana. Journal of Engineering, Design and Technology. 2017; 15(6): 754-77.
[24] Kwak YH, Ingall L. Exploring monte carlo simulation applications for project management. Risk Management. 2007; 9(1):44-57.
[25] Seber GA. Nonlinear regression models. In the linear model and hypothesis 2015 (pp. 117-28). Springer, Cham.
[26] Vyas T, Varia HR. Predicting traffic induced noise using artificial neural network and multiple linear regression approach. International Journal of Advanced Technology and Engineering Exploration. 2022; 9(92): 1009-27.
[27] Stevens JP. Applied multivariate statistics for the social sciences. Routledge; 2012.
[28] Waters D, Waters CD. Quantitative methods for business. Pearson Education; 2008.

Gafel Kareem Aswed received the B.Sc. degree in building and construction engineering and the M.Sc. degree in project management engineering from university of technology (UOT) Bagdad- Iraq. He is a faculty member (Assistant Professor) in Kerbala University-college of engineering, Karbala, Iraq.
Email: ghafil.kareem@uokerbala.edu.iq

Mohammed Neamah Ahmed received the B.Sc. degree in Civil Engineering from university of Babylon, the M.Sc. and PhD degree in project management engineering from university of Bagdad-Iraq. He is a faculty member (Assistant Professor) in Kerbala University-college of engineering, Karbala, Iraq.
Email: mohammed.niema@uokerbala.edu.iq

Hussein Ali Mohammed received the B.Sc. degree in building and construction engineering, the M.Sc. degree in project management engineering from university of technology (UOT) Bagdad- Iraq and the PhD degree from university of Bagdad. He is a faculty member (Assistant Professor) in Kerbala University-college of engineering, Karbala, Iraq.
Email: hussein.a@uokerbala.edu.iq

Appendix I

S. No.	Abbreviation	Description
1	AA	Accuracy Percentage Average
2	ANN	Artificial neural network model
3	AS	Asphalt
4	B	Barriers Cost
5	BTC	Brimelow's model of time–cost
6	C	Final Total Cost
7	CI	Confidence Interval
8	CL	Culverts
9	D	Actual Duration
10	df	Degree of Freedom
11	E	Earthwork
12	FANN	Artificial Neural Network Tool
13	FD	Fuzzy Delphi
14	L	Length of the Road
15	INDOT	Indiana Department of Transportation
16	MAPE	Mean Absolute Percentage Error
17	MLRM	Multi Linear Regression Model
18	NLRM	Nonlinear Regression Model
19	PCA	Principal Component Analysis
21	PDF	Probability Distribution Function
22	R	Rock
23	R2	Coefficient of Determination
24	RMSE	Root Mean Square Error
25	S	Subbase
26	Sig.	Significance Value
27	SPSS	Statistical Package for Social Science
28	VIF	Variance Inflation Factor
29	W	Width
30	WEKA	Waikato Environment for Knowledge Analysis

Appendix II Actual project data used for models' evaluation

Project No.	Length(m)	Width (m)	Duration (days)	Total cost (IQD)	Barriers (IQD)	Rock (m^3)	Earthwork (m^2)	Culverts (m.l)	Subbase (m^2)	Asphelt (m^2)
1	2300	6	210	619158000	15605000	300	18000	5	8500	14500
2	1800	6	180	405715000	9558000	600	7000	0	3510	10800
3	1100	5	180	297330000	1.13E+08	0	6000	132	2145	6600
4	2500	6	270	783589250	16775000	3000	20000	20	6750	15000
5	650	6	120	221550000	58589250	0	5500	75	1600	3900
6	2400	6	300	669439500	32500000	1600	15500	0	4700	14400
7	1800	5	240	416000000	500000	0	12000	8	3250	9000
8	1100	5	180	275260250	80489500	3500	52600	0	2350	5500
9	1100	6	180	427155000	23700000	0	7000	30	2145	6600
10	2450	5	240	665800000	16097750	800	10000	77	5200	12250
11	2400	6	300	669439500	1.53E+08	0	15500	48	4700	14400
12	1800	5	240	416000000	500000	0	12000	8	3250	9000
13	1100	5	180	275260250	80489500	3500	8250	0	2350	5500
14	1100	6	180	427155000	23700000	0	7000	30	2145	6600