

FUNCTIONAL DURATION MODELS FOR HIGHWAY CONSTRUCTION PROJECTS IN NIGERIA

Waziri, Baba Shehu*, Jibrin, Ahmad Tijjani and Kadai, Bukar

(Department of Civil and Water Resource Engineering, University of Maiduguri, Nigeria)

Corresponding author's e-mail address: shehuwaziri@gmail.com

Abstract

Construction duration significantly influences funding, financing and resources allocation decisions that take place early in project design development. This study attempts to develop through regression analysis highway construction duration models by incorporating relevant predictor variables having statistically significant relationship with highway completion time. Historical data of highway projects initiated and completed between 2007 and 2012 were considered to enable the collection of homogenous data in terms of time, cost and other economic variables. Three multiple regression models were developed in the form of linear, semi-log and log-log transformations. The performances of the models were established by measuring their prediction accuracy and goodness of fit over a test sample of 15 successful projects. Results of the analysis showed that all the three models are statistically significant and have good fit to the data with R^2 values of 0.816, 0.877 and 0.970 respectively. The log-log model outperformed the other models with an average percentage error of -2.94%, maximum error of 15.5% and mean absolute percent error (MAPE) of 6.95%. The compare favourably with past studies which have shown that traditional methods of duration estimation at early project stages have values of MAPE typically in the order 10-20%.

Keywords: Highway Construction, Duration Models, Regression Analysis, Nigeria

1. Introduction

Time and cost are extremely important parameters of construction projects essential for planning, feasibility analysis, budget allocation decisions, project monitoring and litigation which often becomes the basis for ascertaining other estimates (Bowen *et al.* 2000; Chan and Chan, 2002; Ogunsemi and Jagboro, 2006; Skitmore and Ng 2003). Shr and Chen (2006) considered time and cost as critical benchmark for evaluating the performance of construction projects which enables effective budgeting and adequate financing. They are also used as an indicator of contractors efficiency, professionalism and competence which reflects the ability of the contractor to organize and control site operation, to appropriately allocate resources and to manage the flow of information to and from the design team and among all other project parties (Ireland, 1986; CIDA, 1993). Determining optimum durations for projects at the early phase is a veritable exercise for incorporating realistic schedule in the bid package as the indirect cost associated with capital projects is highly dependable on the duration of the project.

The prediction of construction duration represents a problem of continual concern to both state highways and contractors since highway construction duration is not always a simple problem because of myriad of factors influencing its accuracy such as project size, type, location, year, types of materials used and method of construction adopted (Hegazy and Ayed, 1998).

However, Mbacu and Olaoye (1998), Izam and Bustani (1999) and Olupoppola *et al.* (2010) have shown that construction delay is a major factor bedevilling the Nigerian construction industry because almost all projects are completed at durations much longer than their initially planned duration. Odusami and Olusanya (2000) reported that construction projects in Lagos experience about 51% delay which could be attributed to the numerous complex factors influencing the construction process. Therefore, it is imperative to develop at the early phase the probable completion duration of highway projects to enable practical estimation and feasibility study.

2. Statement of Research Problem

Despite the importance of accurately and reliably predicting construction duration, early in the project development lifecycle, few tools, practices and procedures are presently available for application (Williams, 2008). The building construction industry has explored this area and found results in statistical regression analysis and artificial intelligence. Such analysis has led to the understanding of the components that influence construction durations and their relationships (Bromilow, *et al.*, 1980; Nkado, 1992; Chan and Kumaraswamy, 1995, Chan 1999 and Skitmore and Ng, 2001). The highway construction industry has not demonstrated such results and both clients and contractors in the industry recognized the need for improved conceptual design level estimating practices. Highway construction duration estimates are prepared very early in project design and based on individual experience on similar projects (William, 2008) which are more often than not unrealistic and inaccurate. The alternative methods adopted by most contractors usually require more detailed scheduling tool and assumptions pertaining to project information such as material quantities and activities involved. Such approaches are time consuming and characterised by lack of information at the early phase.

However, the highway construction industry calls for pragmatic models and robust methods for fast and efficient prediction of construction duration at the planning phase. This study therefore attempts to develop a functional model based on historical data for the prediction of highway project duration by incorporating relevant predictor variables.

3. Highway Duration Models

Duration models are generally developed for reliable prediction of construction durations based on available information at a particular phase of project. Construction duration is estimated either based on clients time constraints or detailed analysis of work to be done and resources available using estimates of time requirement for each specific activity. Most predictive models are based on historical project data of successful projects prepared by construction professionals for determining project duration which has been a subject of many studies (Bromilow, 1969; Izam and Bustani, 1999; Maghrebhi *et al.*, 2013; Shr and Chen 2006; Levinson and Karamalaputi, 2003).

The first empirical modelling of construction time performance (CTP) was carried out by Bromilow (1969) in Australia. The resulting model is referred to as the Bromilow's Time-Cost Model (BTC) which enables construction duration to be determined based on estimated final cost. Shr and Chen (2006) developed a model to illustrate cost-time relationship of

highway projects using data obtained from the Florida Department of Transportation. The model provides state highway agencies and contractors with increased control and understanding regarding the time value of highway construction projects. However the model was not suitable for projects with great degree of change order. Martin *et al.* (2006) established a forecasting model using regression analysis based on actual time of building construction in the United Kingdom. The result presents a tool to aid clients and contractors in estimating or benchmarking the construction duration at the earliest stage of future projects. Hoffman *et al.* (2007) developed a multiple regression model to predict highway construction duration by improving the parameters that impact construction duration. The data for the study were examined using BTC Model and regression analysis. The regression model compared favourably with the BTC Model with minimum error. Assadulla (2010) also attempted to develop highway duration model using stepwise regression and Artificial Neural Network. The outcome of the study revealed that the ANN represented higher accuracy and reliability. Petrusseva *et al.* (2013) conducted a research and developed a forecasting model for construction time using support vector machine. The analysis involved 75 objects structured in the period 1999-2011 in the Federation of Bosnia and Herzegovina. The results showed an accurate presentation for building work and suggested it to be useful for planning in the construction industry. Waziri and Yusuf (2014) made an attempt to validate the BTC model using highway construction data in Nigeria. The model indicated a K value (which demonstrates a general level of time performance) of 2.8 which is considered very low compared to values obtained by previous studies for other categories of construction. The model also showed a weak prediction efficacy with MAPE of 19% over a test sample. This may be attributed to the non inclusion of other duration influencing factors apart from cost in the model. Therefore it is necessary to consider other relevant variables in order to accurately and reliably estimate the probable completion time of highway project for the purpose of feasibility studies, planning, budgeting and adequate financing.

4. Methodology

4.1 Data Collection

Data for analysis and modelling were obtained from the records of completed highway projects from clients and contractors in Gombe and Bauchi states in North Eastern Nigeria. Dataset of 57 successful highway projects initiated and completed between 2007 and 2014 were considered to enable the collection of homogenous data in terms of time and cost. Projects that made up the survey population were restricted to those with contract value of more than ₦100,000,000 which are considered to demonstrate reasonable scope and complexity (Waziri and Yusuf 2014). 75% of the projects (42) were used for developing the models while the remaining 25% (15) were used for evaluation and validation of the models. Three regression analyses viz; linear, semi-log, and log-log were established incorporating four (4) significant variables affecting highway construction duration with the view to improving the BTC model which considers only the cost of construction. The variables considered and included in the models as input variables are:

- i. Estimated project cost per unit length

- ii. Number of culverts along the highway stretch
- iii. Thickness of pavement materials
- iv. Road length

Linear regression analysis was employed to describe the value of dependent variable on the basis of the independent variables. The linear regression equation is as presented in equation (i)

$$y_i = \beta_o + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 \quad (1)$$

Semi-log regression analysis was also employed for developing a transformed regression model. It is identified to produce best statistical results in terms of parameter significance (Stoy and Schalcher, 2007). Semi-Log regression is of the form presented in equation (2)

$$\ln y = \beta_o + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 \quad (2)$$

Log-log model having all the variables transformed to the logarithmic form was also developed; it is of the form presented in equation (3)

$$\ln y = \ln\beta_o + \beta_1\ln x_1 + \beta_2\ln x_2 + \beta_3\ln x_3 + \beta_4\ln x_4 \quad (3)$$

The models were then evaluated for goodness of fit and prediction performance by comparing the prediction results over a test sample using Mean Absolute Percent Error (MAPE) and Mean Square Error (MSE). The equations for the error measures are as presented by equation (4) and (5) respectively

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\text{actual}_i - \text{predicted}_i}{\text{predicted}_i} \right| \times 100 \quad (4)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (\text{actual}_i - \text{predicted}_i)^2 \quad (5)$$

5. Results and Discussion

5.1 Relationship of Input Variables

The input variables were examined to determine the existence of autocorrelation between them. The product moment correlation which illustrates the relationship between them was computed and presented in Table 1.

Table 1: Pearson product moment correlations between each pair of input variables

Variables	Correlation Coefficient	P value
Duration – Road Length	0.062	0.042
Duration – No of Culverts	0.809	0.000
Duration – Road Thickness	-0.336	0.025
Duration – Cost/unit length	0.112	0.045
Road Length – No. of Culverts	0.116	0.023
Road Length – Road Thickness	-0.066	0.279
Road Length – Cost/Unit length	0.726	0.000
No. of Culverts – Road thickness	-0.309	0.246
No. of Culverts – Cost/unit Length	0.148	0.048
Road Thickness – Cost/ unit Length	0.117	0.046

The product moment correlation between each pair of the variables measures the relationship between the variables. These coefficients range between -1 and +1. P-values below 0.05 indicate statistically significant non-zero correlations at the 95% confidence level. From the result, the pairs of variables having P-value greater than 0.05 are road length and road thickness and No. of culverts and road thickness with values 0.279 and 0.246 respectively indicating zero correlation at the 95% confidence level.

5.2 Regression Models

Three different regression models namely linear, semi-log and log-log were developed. The semi-log has the dependent variables transformed to the logarithmic form while in the log-log model all the variables were transformed into their respective logarithmic form.

5.2.1 Linear Regression Model (M1)

The summary of the multiple regression model in which all the variables were considered without transformation is presented in Table 2 while the equation of the fitted model is presented in equation (6).

Table 2: Summary of Linear Regression analysis

R	R ²	Adj R ²	Std Error	Sig. F Change	P Value	Durbin-Watson
0.816	0.666	0.630	333.09	18.43	0.000	1.156

From the results, the R² value of 0.666 indicates that 66.6% of the variability in construction duration is explained by the model indicating significant relationship between the dependent and the independent variables. Since the p value is also less than 0.005 it indicates that there is a statistically significant relationship between the variables at the 95% confidence level. The parameter estimate of the independent variables is presented in Table 3.

Table 3: Variables parameters for Linear Model

Variable	Coefficient	Std Error	t value	P value
Constant	374.047	392.802		0.347
Road length	-0.062	0.042		0.532
No. of Culverts	31.447	4.152		0.000
Road thickness	-0.037	0.349		0.292
Cost/unit length	0.377	0.726		0.607

The results of the coefficients revealed that the p value of all the three variables except number of culverts are greater than 0.05 indicating weak association between the independent variables and construction duration as presented by the model. The equation of the fitted model is presented in equation 1.

$$Duration = 374.047 - 0.026 * Road Length + 31.447 * No. of Culverts - 0.373 * Road Thickness + 0.377 * Cost per unit length \quad (6)$$

5.2.2 Semi-Log Regression

In the semi-log regression analysis, the dependent variable of the regression analysis was transformed to its logarithmic form in order to improve the performance of the model. The data for the duration is transformed while the data of the independent variables were entered raw. The summary of the regression is presented in Table 4.

Table 4: Summary of the Semi log Regression

R	R ²	Adj R ²	Std Error	Sig. F Change	P Value	Durbin-Watson
0.877	0.768	0.743	0.536	30.69	0.000	1.665

From the result of the analysis, the R² value of 0.768 indicates that 76% of the variability in the data is accounted for by the model. The R value of 0.877 shows significant association between the dependent variable and the independent variables. The p value of <0.001 suggests significant relationship between the variables at the 95% confidence level. The parameter of the variables of the semi-log regression model is presented in Table 5.

Table 5: Variables parameters for Linear Model

Variable	Coefficient	Std Error	t- value	P value
Constant	5.182	0.625	8.296	0.000
Road length	-0.00000737	0.000	-1.112	0.274
No. of Culverts	0.062	0.007	9.438	0.000
Road thickness	-0.001	0.001	-1.403	0.169
Cost/unit length	0.002	0.001	1.928	0.052

The results of the coefficients of the variables indicate that the p value corresponding to the variables road length and road thickness are greater than 0.005 suggesting the weak contribution of the variables to the model. This may be attributed to the fact that some of the properties contained in those variables may be contained in some of the variables showing strong association. The semi-log model is presented is represented by equation (7).

$$\ln \text{Duration} = 5.182 - 7.368^{-005} * \text{Road Length} + 0.62 * \text{No. of Culverts} - 0.001 * \text{Road Thickness} + 0.002 * \text{Cost per unit length} \quad (ii) \quad (7)$$

5.2.3 Log-Log Regression

In order to further test for the possible improvement of the earlier models, both dependent and the independent variables were transformed to the logarithmic form to have the log-log regression model. The summary of the analysis is presented in Table 6.

Table 6: Summary of Log-Log Regression

R	R ²	Adj R ²	Std Error	R Square Change	Sig. F Change	P Value	Durbin-Watson
0.970	0.941	0.934	0.271	0.941	143.09	0.000	1.225

From the analysis, 94% of the variability in duration is explained by the model as indicated by the R² value of 0.941. The adjusted R-squared statistics which is more suitable for comparing models with different numbers of independent variables is 93.4%. The standard error of the estimate shows the standard deviation of the residuals to be 0.271. This value can be used to construct prediction limits for new observations. The p value of <0.001 indicates that the relationship is significant at 99% confidence level. Durbin-Watson statistics of 1.225 is use to test the residuals to determine if there is any significant correlation in which the data occur. Since the p value is less than 0.005, there is indication of serial autocorrelation in the residuals. The result of the parameters for the variable in the log-log model is presented in Table 7.

Table 7: Parameters estimates of Variables for Log-Log Model

Variable	Coefficient	Std Error	t-value	P value
Constant	4.514	2.351	1.920	0.043
Road length	-0.005	0.069	-0.068	0.446
Road thickness	-0.271	0.320	-0.846	0.043
No. of Culverts	1.260	0.063	19.923	0.000
Cost/unit length	-0.031	0.070	-0.438	0.064

From the results of the parameters of the independent variables, the highest p value on the independent variables is 0.446 belonging to road length. This is followed by 0.064 belonging to cost/unit length. Since the p values corresponding to these variables are greater than 0.05, these terms are not significant at the 95% confidence level. The equation of the model is presented in equation (8).

$$\ln \text{Duration} = 4.514 - 0.005 * \ln \text{Road Length} + 1.260 * \ln \text{No of Culverts} - 0.27 * \ln \text{Road Thickness} - 0.031 * \ln \text{Cost per unit length} \quad (8)$$

The summary of the three models are presented in Table 8

Table 8: Summary of the models

Model	R	R ²	Adj R ²	STD Error	Sig F change	P value
Linear	0.816	0.666	0.630	337.09	18.43	<0.000
Semi-log	0.877	0.768	0.743	0.536	30.69	<0.0001
Logarithmic	0.970	0.941	0.934	0.271	143.09	<0.0001

5.3 Evaluation of Models

5.3.1 Prediction Performance

The prediction performance of the models was established by comparing their predictions over a test sample. The sample consists of dataset of 15 successfully completed highway projects. The model equations were constructed using Microsoft excel spreadsheet and were used to predict the construction durations of the test sample. The result is presented in Table 9.

Table 9: Prediction Performance of Models

Model	Average Error (%)	Min Error (%)	Max Error (%)	MAPE (%)
Linear	-5.37538	-96.2595	137.3687	52.63
Semi-log	-91.6635	-99.6348	-84.3649	63.66
Logarithmic	-2.94126	-12.8998	15.50079	6.95

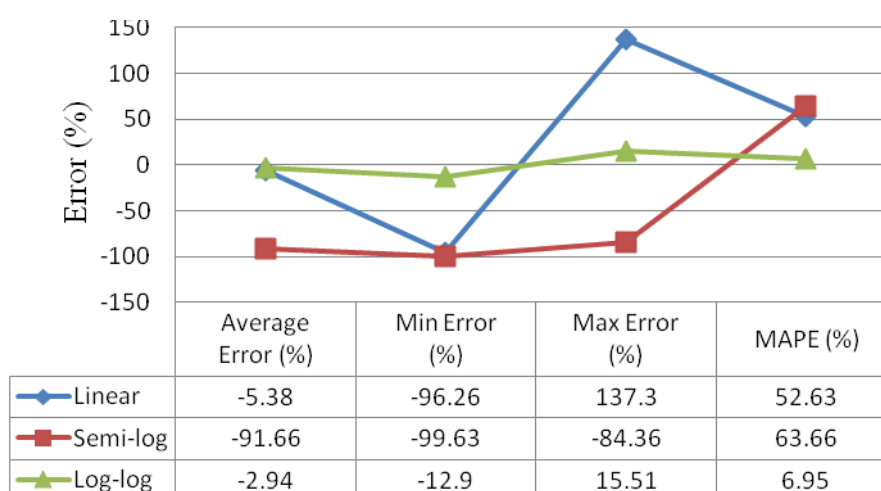


Figure 1: Predictive performance of duration models

The results of the evaluation revealed that the linear, semi-log and log-log models have average prediction errors of -5.38, -91.66 and -2.94 % respectively which suggests that the

log-log model has the smallest average error over the test sample. The result also revealed that the log-log model has the smallest value of MAPE of 6.95% demonstrating a good performance. The logarithmic model therefore outperformed the linear and the semi-log models in all the error terms.

5.3.2 Prediction Accuracy of Log-Log Model

The calculated output of the test samples by the log-log model revealed that 73.33% of the test samples were underestimated while 26.67% were overestimated. The range of underestimating varies from -10.88% to -1.03% with an average value of -7.23%. The range of overestimating varies from 0.10% to 15.50% with an average value of 7.51%. These ranges show that the model is skewed to underestimating the test samples. On the overall, the log-log model has an average % error of -2.94%, a maximum % error of 15.51% and a MAPE of 6.94%. The value of MAPE of 6.94% is a very good performance; it is within the acceptable range of $\pm 10\%$ by Ogunsemi and Jagboro (2006), $\pm 25\%$ for early estimates by Schexnayder and Mayo, (2003).

6. Conclusion

Reliable and accurate estimates of highway project durations are desired by state highway agencies and contractors for inclusion in the bid package. In the study, four relevant predictor variables namely road length, road thickness, number of culverts and cost per unit length were included in the model in order to improve the performance of the models. This is because earlier attempts to modelled using only cost of construction as the independent variable in line with BTC model was found to be unsatisfactory. The results of the study revealed 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. This may be attributed to the transformation of the variables to the logarithmic form. The model is useful to both clients and contractors because of its ability to predict construction duration at the early project stage, since the information necessary as input variables can be extracted from scope designs and early estimates. Due to its simplicity, it can be handled using Microsoft excel or a simple computer programme. The main limitation of regression models is that the collinearity of the predictor variables may affect the prediction performance of the regression model.

References

- Asadullah, A. 2010. Development of Neural Network Models for Prediction of Highway Construction Cost and Project Duration. An Honors Thesis presented to the Faculty of The Russ College of Engineering and Technology, Ohio State University Columbus, OH43210.
- Bowen, PA., Hall, KA. Edwards, PJ., Pearl, RG. and Cattell, 2000. Perceptions of time, cost and quality management of building projects. *The Australian Journal of Construction Economics and Buildings*, 2(2): 48-56.
- Bromilow, FJ. 1969. Contract time performance expectations and the reality. *Building Forum*. 1(3):70-80.
- Bromilow, FJ., Hinds, MF., and Moody, NF. 1980. AIQS Survey of building contract time performance. *Building Economist*, 19(2): 79-82.

Chan, APC. 1999. Modelling building durations in Hong Kong. *Construction Management and Economics*, 17:189-196.

Chan, APC and Chan, DWM. 2002. Benchmarking project construction time performance: the case of Hong Kong, Project management-impresario of the construction industry held 22nd-23rd March 2002, in Hong Kong.

Chan, DWM. and Kumaraswamy, MM. 1995. A study of factors affecting construction durations in Hong Kong. *Construction Management and Economics*, 13:319-333.

CIDA. 1993. Project performance update – A report on the cost and time performance of Australian building projects completed 1988-1993, Construction Industry Development Agency, Australia.

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

Hoffman, GJ., Thal, AE., Web, TS. and Weir, JD. 2007. Estimating performance time for construction of projects. *Journal of Management in Engineering*, 23(4):193-199.

Ireland, V. 1986. A comparison of Australia and US building performance for high-rise buildings. University of Technology, Sydney, Australia: School of Building Studies.

Izam, YD. and Bustani, SA. 1999. Predictive duration models for building construction in Nigeria. *Journal of Environmental Sciences*, 3(2): 125-131.

Levinson, D. and Karamalputi, K. 2003. Predicting the construction of new highway links. *Journal of Transportation and Statistics*, 6(1):2-9.

Maghrebi, M., Afshar, A. and Maghrebi, MJ. 2013. A novel mathematical model for determining time-cost trade-off based on path constraints. *International Journal of Construction Engineering and Management*, 2(5):137-142.

Martins, J., Burrows, JK. and Pegg, I. 2006. Predicting construction durations of buildings projects. XXIII FIG congress, Munich, Germany.

Mbachu, JIC. and Olaoye, GS. 1999. Analysis of major delay factors in building projects execution, *Nigerian Journal of Construction Technology and Management*, 2(1):81-86.

Nkado, RN. 1992. Construction time information system for the building Industry. *Construction Management and Economics*, 10:489-509.

Olupoppola, FO., Emanuel, OO., Adeniyi, M.F. and Kamaldeen, AA. 2010. Factors affecting time performance of buildings. Proceedings of the Construction, Building and Real Estate Research Conference of the Royal Institution of Chartered Surveyors, Dauphine, Paris.

Ogunsemi, DR. and Jagboro, GO. 2006. Time-cost model for building projects in Nigeria. *Construction management and economics*, 24:253-258.

Odusami, KT. and Olusanya, OO. 2000. Client's contribution to delays on building completion, *The Quantity Surveyor*, 30(1): 30-33.

Petruseva, S., Zicska-panakovska, V. and Zudo, V. 2013. Predicting construction duration with support vector machine. *International Journal of Research in Engineering and Technology*, 2(11):12-24.

Schexnayder, CJ. and Mayo, RE. 2003. *Construction Management Fundamentals*, McGraw-Hill Higher Education, Boston, MA.

Shr, F. and Chen WT. 2006. Functional model of cost and time for highway construction projects. *Journal of Marine Science and Technology*, 14(3): 127-138.

Skitmore, RM. and Ng, ST. 2001. Australian project time-cost analysis: Statistical analysis of Inter-temporal trends. *Construction Management and Economics*, 19:455- 458.

Skitmore, RM. and Ng, ST. 2003. Forecast models for actual construction time and cost. *Building and Environment*, 8(8):1075-1083.

Stoy, C. and Schalcher, HR. 2007. Residential building cost drivers and indicators. *Journal of Construction Engineering and Management*, 133(2):139-145.

Waziri, BS. and Yusuf, N. 2014. Application of Bromilow's time-cost model for highway projects in Nigeria. *Arid Zone Journal of Engineering Technology and Environment*. 10: 1-11.

Williams, R. (2008). *Development of Mathematical Model for Preliminary Predictions of Highway Construction Duration*. P.hD. Dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA.