Journal of Construction in Developing Countries, 29(Supp. 1), 107–128, 2024

Modelling Cost-Risk Impact on Nigerian Highway Projects Using Multiple Linear Regression and Artificial Neural Networks

Victor Imohnikhe Aligamhe[1], *Muzani Mustapa[1], Chukwuemeka Patrick Ogbu[2] and Mohd Saidin Misnan[1]

Published: 15 October 2024

To cite this article: Victor Imohnikhe Aligamhe, Muzani Mustapa, Chukwuemeka Patrick Ogbu and Mohd Saidin Misnan (2024). Modelling cost-risk impact on Nigerian highway projects using multiple linear regression and artificial neural networks. *Journal of Construction in Developing Countries*, 29(Supp. 1): 107–128. https://doi.org/10.21315/jcdc.2024.29.S1.6

To link to this article: https://doi.org/10.21315/jcdc.2024.29.S1.6

Abstract: Cost predictive models for highway projects are relatively scarce in developing countries, despite the frequency and magnitude of project cost overruns in such countries. This study identified critical cost risks impacting highway projects and modelled their impacts on the actual cost of the projects. Historical cost data on highway projects published by the Nigerian Federal Ministry of Power, Works and Housing in 2017 served as a preliminary list of projects for the study, while further cost data were obtained from highway engineers and quantity surveyors across Nigeria using the snowballing technique until 103 highway projects were identified. Project participants were purposively chosen to fill out questionnaires on cost-risks factors associated with highway construction projects. The relative importance index and Pareto 80/20 rule were used to analyse the collected primary data. Thereafter, multiple linear regression and artificial neural network models were developed. Findings revealed that the increase in the cost of construction materials and labour and the fluctuations in foreign exchange rates were the most significant risks impacting highway project cost performance. A comparison of the models indicated that the artificial neural networks model performed better. Hence, the artificial neural networks model is a superior technique for modelling the relationship between cost risks and cost performance of highway projects.

Keywords: Nigerian highway projects, Artificial neural networks, Multiple linear regression, Cost risk of highway projects, Cost performance of highway projects

INTRODUCTION

Highway infrastructure is critical to the economic development of any nation, especially developing nations. According to Hamma-Adama et al. (2021), modern roads are crucial in context due to their ability to stimulate social and economic benefits. Because of the near absence of other means of transportation in developing countries like Nigeria, over 90% of people, goods and services are conveyed by roads (Anigbogu, Ahmad and Molwus, 2019). However, diverse factors have hampered the actualisation of an effective and efficient highway infrastructure due to uncertainties encountered during the project planning and execution phases. According to Okate and Kakade (2019), highway construction projects have enormous risks caused by their prevailing underground conditions and extensive geographical and regional spread.

[1]Faculty of Built Environment and Surveying, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, MALAYSIA
[2]Faculty of Environmental Sciences, University of Benin, Ugbowo Campus, PMB 1154, Edo State, NIGERIA
*Corresponding author: muzani@utm.my

The initial and final costs of infrastructural projects often experience notable deviations due to underestimation at their planning stage (Kovacevic et al., 2021). Construction project cost overruns have remained a global problem. For example, Changali, Mohammad and Van Nieuwland (2015) claim that 98% of megaprojects experience cost and schedule overruns. Love, Edwards and Iran (2012) report that schedule and cost overruns could occasionally amount to 183% and 70% more than the initial stipulated estimate. According to Tariq and Gardezi (2023), the primary reasons for cost overruns are inadequate project planning and management, poor participant communication, poor management of materials, failure of equipment, variability in project solutions, poor management of resources and the use of outdated technologies and tools. According to Al-Zwainy and Aidan (2017) and Sodikov (2005), the absence of preliminary data, the absence of a database of road work costing and the lack of modern cost estimation techniques are the main challenges encountered when conducting cost estimation for road projects at the planning phase. Additional challenges are caused by uncertainties resulting from engineering solutions, environmental issues and socio-economic issues (Sodikov, 2005).

With the cost of constructing roads and maintaining the road infrastructure growing steadily every year, government and highway authorities need to identify and analyse the impact of risks on the cost performance of such construction projects. With the shortcomings of traditional estimation methods, which usually fail to consider the impact of risks on project performance, researchers and professionals in the construction industry are gradually embracing modern estimation methods. Artificial neural networks are among the modern approaches that have been recently employed in construction cost estimation. Notwithstanding the presumed superiority of modern modelling techniques over traditional models, there is a need for a comparison of models' results using the same data parameters in order to ascertain this presumption. Based on the foregoing, it is, therefore, imperative to identify and quantify the impact of risks on cost performance in Nigerian federal highway projects, especially since the findings from a systematic review conducted by Awuku et al. (2022) showed that North America, Asia, Europe and the Middle East contributed the most to enhancing highway cost estimation research between 1983 and 2019.

Challenges in Nigeria's federal highway projects centre on cost estimation and management (Ikechukwu and Akiohnbare, 2017). There is a notable gap in research regarding the effectiveness of artificial neural networks in modelling cost risk, comparing them with traditional methods, quantifying risk impact and incorporating artificial neural network-based models into current practices. This study sought to address these gaps to enhance the effectiveness of highway infrastructure development. This paper aimed to develop multiple linear regression and artificial neural network models for predicting the conceptual cost of highway projects in Nigeria. The research objectives were as follows:

1. Identify the cost-risk factors that affect the performance of Nigeria's federal highways.
2. Use multiple linear regression and artificial neural networks to forecast the impact of cost risks on Nigeria's federal highway projects.
3. Validate the models that have been developed and compare their performance.

LITERATURE REVIEW

Using conventional techniques, Leo-Olagbaye and Odeyinka (2020) developed impact risk predictive models for highway projects in Nigeria using multiple linear regression. The findings, based on 37 historical highway projects, showed that 39.7% of the variation in cost performance could be explained by five variables. For example, the coefficient of correlation (R^2) = 0.397 implies that a model lacks the needed accuracy for predicting the actual cost of road projects in Nigeria. Highway project costs are impacted by several risk factors, but a few of these key risk factors are usually considered when using conventional modelling estimation techniques due to their inability to deal with multicollinearity among risk factors.

Sodikov (2005) developed a cost estimation model for road projects in developing countries using the artificial neural networks technique, particularly in Poland and Thailand, due to their relatively large number of projects. The study explored the relationship between the project estimate and other variables such as work activities, terrain types and road parameters. Findings revealed that the width of the pavement, volume of earthwork and the duration of work have a high influence on the estimate of a new road project. Al-Suhaili, Saco and Al-Zwainy (2010) used the artificial neural networks technique to develop a road project cost estimation model. The developed model exhibited a good degree of accuracy, with an R^2 of 84.95%. However, it is important to note that the model did not consider the impact of risk on cost estimation. Al-Zwainy and Aidan (2017) developed an artificial neural network cost prediction model for the structural work of highway projects in their planning phase. Factors affecting the cost parameter predictions were given, and equations were used to estimate the cost of the structural works for the highway project. The study concludes that the artificial neural networks model could predict the cost of structural work for highway projects with a high degree of accuracy (93.19%). El-Kholy (2019) used four artificial neural network-based paradigms as the principal component analysis in predicting delay and cost overrun percentages (PDCOP) for highway projects in Egypt. The study adopted 15 cost overrun factors as predictors. Findings show that the artificial neural network-based paradigm improved the prediction model's accuracy and reliability for percentage delay and cost overrun. Fernando, Dishan and Zhang (2023) developed artificial neural network models that focused on predicting the costs of significant elements of bridge construction projects such as piling, piers, abutments, pre-stressed beams, concrete slabs, bridge paving, bridge furniture and miscellaneous. However, this study did not consider the impact of risk on the cost performance of infrastructural projects, therefore assuming that infrastructural projects were not impacted by risk.

Although studies have previously been carried out by researchers exploring artificial neural network techniques in cost estimation for highway projects in developing nations (e.g., Al-Zwainy and Aidan, 2017; Sodikov, 2005; Al-Suhaili, Saco and Al-Zwainy, 2010; El-Kholy, 2019), it was observed that most of the studies did not consider the impact of risk on project actual costs and that there was no record of similar studies conducted in Nigeria. Therefore, this study was essential due to the limited research on Nigerian highway projects using modern techniques in conceptual cost estimation capable of evaluating the impact of risk on the project's final cost. Considering that Nigeria is the world's most populous black nation, covering 923,768 square kilometres, it is therefore imperative to develop and implement accurate cost prediction models for Nigeria's construction industry, particularly its highway projects.

Multiple Linear Regression

Multiple linear regression is a straightforward and understandable model best suited for linear relationships. A regression model calculates constant values that reflect how changes in independent variables affect the dependent ones. Multiple linear regression, as a statistical tool, elucidates the relationship between dependent variables (such as highway project cost overruns) and independent variables (e.g., cost-risk factors). The first step involved selecting relevant risk variables based on literature and expert input and then identifying and gathering data on these chosen variables.

Artificial Neural Networks

Based on the adjustment of weights, artificial neural networks enable learning through training and generalising the behaviour of a problem (Barros, Marcy and Carvalho, 2018; Pineda-Jaramillo, Insa and Martínez, 2017). It uses mathematical algorithms based on statistical models to identify data trends to provide descriptions or predictions (Sarker, 2021). The artificial neural network is a powerful and versatile model that can capture complex patterns and non-linear interactions. It can forecast the cost of construction projects with superior accuracy compared to other AI-based algorithms due to their ability to learn from historical data through training to generalise results (Glymis et al., 2017; Tijanić, Car-Pušić and Šperac, 2017).

Artificial neural networks is a method of computation that aids decision-making and can address most of the significant shortcomings of conventional estimating methodologies, thereby reducing economic risks by automatically analysing vast volumes of project cost data and producing precise estimates (Awuku et al., 2022; Elbeltagi et al., 2014). Notwithstanding the presumed superiority of modern modelling techniques over traditional models, there is a need for a comparison of the model's results using the same data parameters to ascertain this presumption.

Prior literature review underscores challenges in Nigerian highway infrastructure projects and primarily centres on cost estimation and management. However, the research gap lies in the scarcity of studies examining the efficacy of artificial neural networks as a cost-risk modelling technique for these projects. There is a dearth of empirical evidence showcasing artificial neural network applications in Nigerian federal highway projects, which is in contrast with its recognition in developed countries. Secondly, there is a lack of comparative studies directly pitting traditional methods against artificial neural networks within the Nigerian context, hindering insights into artificial neural networks' suitability. Furthermore, while risks' impact on project costs is acknowledged, there has been insufficient research quantifying this impact using artificial neural networks in Nigerian highway projects. Finally, there's an absence of exploration of integrating artificial neural network-based cost-risk models into existing risk management practices in Nigeria. Addressing these gaps promises advancements in construction cost estimation, risk management knowledge and practical guidance for enhancing highway infrastructure development efficiency in Nigeria.

METHODOLOGY

This section discussed the following subsections: Identification of Cost-Risk Factors, Data Collection and Data Analysis.

Qualitative Research: Identification of Cost-Risk Factors

Factors impacting highway performance have been investigated in previous studies, resulting in the identification of 140 highway risks. Consequently, further examination and processing should be applied to extract cost-related risks from the identified total highway risk. A total of 10 experts in highway construction, consultancy and management with 15 years to 25 years of experience (Glymis et al., 2017; Gondia et al., 2019), were chosen from the industry and academics to partake in a focus group discussion (FGD).

Content analysis was conducted on the responses received from the FGD. The results of the content analysis indicated 29 cost-risk factors. Finally, the outcome of the FGD and content analysis served as constructs presented in the questionnaire to enable further investigation.

Quantitative Research: Survey and Statistical Analysis

This section consists of the quantitative approach adopted to achieve a significant part of the study's objectives.

Data Collection

Cost data for highway projects was partly collected from a publication in 2017 by Nigeria's Federal Ministry of Works and Housing. A total of 229 highway projects made up the total number of projects on the list. Despite examining a total of 68 completed projects, including those finished by January 2022, the information for five projects was unable to be accessed. As a result, only 63 projects from this source were included in the study. Recently completed projects were identified for inclusion through a pilot study, resulting in 40 highway projects totalling 103 (refer to Appendix).

During the pilot study, researchers utilised the snowballing technique to contact quantity surveyors and highway/civil engineers across the six geopolitical zones. Respondents were provided with a pro forma to supply the required information. Questionnaires were then distributed to construction professionals, including quantity surveyors and highway/civil engineers, both in hardcopy and electronically. These professionals, who participated in the projects as clients, contractors, or consultants, were chosen due to their frequent roles as construction cost and project managers during highway project implementation. Table 1 presents the demographics of the study respondents.

Table 1. Demographics of respondents

Respondent's Information	Category	Frequency (*n*)	%
Post-qualification experience (years)	Less than five years	9	8.70
	6 years to 10 years	19	18.40
	11 years to 15 years	43	41.70
	16 years to 20 years	27	26.20
	More than 20 years	5	4.90
	Total	103	100.00
Occupational category	Client	49	47.57
	Consultant	26	25.25
	Contractor	28	27.18
	Total	103	100.00
Number of projects in the geographical locations	Southwest	19	18.45
	Southeast	17	16.50
	South-South	18	17.48
	Northeast	17	16.50
	Northwest	13	12.62
	Northcentral	19	18.45
	Total	103	100.00

DATA ANALYSIS

This section consists of the following subsections: Cost Performance of Highway Projects, Significant Cost-Risk of Highway Projects, Multiple Linear Regression Model Development and Artificial Neural Networks Model Development.

Cost Performance (Experiment Data) of Highway Projects

The historical cost data and its analyses in developing the models are given in the Appendix. The percentage cost overruns represent the deviation from the estimated cost of the contract and the actual costs (as shown in Equation 1). To arrive at a comparable digit, a five-point Likert scale was used to measure the severity of the risk impact. The percentage cost overruns were divided by 100.

$$P_{co} = \times 100 \qquad \text{Eq. 1}$$

where P_{co} = Percentage cost overrun, AEC = Actual construction cost and ECC = Estimated construction cost.

Determination of Significant Cost-Risk Factors

This section discusses the methods of data analysis required to determine the significant cost-risk factors of Nigeria's federal highway projects in the following subsections: Relative Importance Index (RII) and Pareto 80/20 Rule.

Relative importance index

The responses of respondents were obtained and evaluated using the RII approach, which has been previously used in other construction management studies (El-Sayegh and Mansour, 2015; Thaseena and Vishnu, 2017). The responses were based on a Likert scale of 1 to 5, as illustrated in Table 2. The RII scores of each risk factor were also classified into a five-point category from the risk factor with the highest RII downwards.

Table 2. Likert scale and RII classification

Points	Likert Scale	RII Classification
1	Very low	< 1.5
2	Low	$1.5 \leq RII < 2.5$
3	Moderate	$2.5 \leq RII < 3.5$
4	High	$3.5 \leq RII < 4.5$
5	Very high	$4.5 \leq RII < 5.00$

Pareto 80/20 rule

This study applied the Pareto principle to determine the most significant variables to achieve a sizable number of significant cost-risk factors. According to the rule, 80% of consequences emerge from 20% of causes (Grosfeld-Nir, Ronen and Kozlovsky, 2007). The Pareto 80/20 rule aims for fewer activities when assessing their total productivity (Rizwan and Iqbal, 2011) and helps certain businesses design rapid models. Based on the rule, the first six factors, according to ranking, as shown in Table 3, were considered critical risks. These factors were then used as independent variables to develop the models.

Table 3. The ranking of cost-risk factors

Rank	Variables	RII	Risk Level	Ref. Code
1	Inflation in the cost of construction materials and labour	4.33*	H	CR9
1	Foreign exchange rate fluctuation/variation	4.33*	H	CR7
3	Changes in input resource prices/Variations in raw material prices	4.32*	H	CR5
4	Improper feasibility study	4.07*	H	CR19
5	Construction cash flow problems/Project funding challenges	3.98*	H	CR3

(*Continued on next page*)

Table 3. *Continued*

Rank	Variables	RII	Risk Level	Ref. Code
6	Unexpected location/Ground conditions	3.90*	H	CR13
7	Change in government/Political change	3.88	H	CR24
8	Interest rate fluctuation/increase	3.83	H	CR8
9	Uncertainty of project budget	3.82	H	CR12
10	Poor communication/coordination between construction parties (owner, consultant and contractor)	3.81	H	CR29
11	Lack of joint risk management mechanism by the contractor and parties	3.75	H	CR22
12	Mismanagement of site and supervision by contractor	3.70	H	CR21
13	Lack of experienced contractor	3.69	H	CR17
13	Lack of experienced consultant	3.69	H	CR18
15	Not applying cost control	3.65	H	CR23
16	Incompetent project supervision/Poor project management	3.63	H	CR15
17	Mishandling of resources	3.42	M	CR20
18	Multiple approval problems	3.28	M	CR28
19	Poor estimating/Inaccurate cost estimate	3.27	M	CR11
20	Discrepancies between actual and contractual quantities	3.25	M	CR27
21	Lack of professionals/experts	3.23	M	CR16
22	Unethical practices/corruption/fraud/bribe	3.22	M	CR25
23	High cost of maintenance	3.18	M	CR14
24	Insufficient design details/specification	3.11	M	CR1
25	Health, safety and environmental (HSE) issues	2.85	M	CR26
26	Low budgeting	2.80	M	CR10
27	Claims	2.79	M	CR6
28	Scope vagueness	2.72	M	CR2
29	Changes in taxation/New tax rates	2.28	L	CR4

Notes: *First six significant risk factors as determined by the Pareto 80/20 rule; H = High, M = Moderate, L = Low.

Data Partitioning and Model Performance Criteria

Prior to the partitioning of the data, outliers were removed, resulting in 97 total experimental datasets (dependent variables as presented in the Appendix). This is to enhance model stability. The data were then categorised into two sets of 80:20 (80% and 20%, respectively). This represents 80% (78 projects) and 20% (19 projects) of the training dataset. The training data were selected randomly from the 97 experimental data points by *R*-statistical software. The process of the multiple linear regression and artificial neural networks model development techniques involved multiple linear regression and artificial neural networks. The artificial neural networks model usually requires optimisation to minimise the prediction errors.

The multiple linear regression and artificial neural network performances were evaluated by assessing their predictive accuracy through the test datasets using the mean absolute percentage error (MAPE), mean square error (MSE) and root mean square error (RMSE). These processes were undertaken to calculate and quantify the error and validate the model.

Multiple linear regression

The critical cost-risk factors constituted the independent variables, while the historical data on cost made up the dependent variables upon which the multiple linear regression model was developed using *R*-statistical software. The regression equation is presented in Equation 2.

$$Yc = a + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \ldots + \beta_nX_n \qquad \text{Eq. 2}$$

Representing the dependent variable as Yc, a = Regression constant, $\beta_{1,}\ \beta_2 \ldots \beta_n$ = Regression estimates, $X_{1,}\ X_2 \ldots X_n$ = Critical cost risk factors (independent variables) as presented in Table 3.

Model development using the multiple linear regression technique

The multiple linear regression model development output is presented in Table 4. One predictor was eliminated during the development of the multiple linear regression model, indicating a high degree of multicollinearity among the variables, as evident in the high variance inflation factor (VIF) and low tolerance values. The models and their coefficients, as presented in Table 4, were substituted into Equation 2, and the mathematical expression of the multiple linear regression model is presented in Equation 3:

$$Yc = 0.551 + 0.011X_1 + 0.029X_2 + 0.026X_3 + 0.032X_4 - 0.035X_5 \qquad \text{Eq. 3}$$

The multiple linear regression model predicts cost overruns (Y) based on cost risk factors acting as independent variables. In Equation 4, Yc represented the predicted cost overrun, while X_1, X_2, X_3, X_4 and X_5 represent different cost risk factors. The coefficients (0.011, –0.029, 0.026, 0.032, –0.035) indicated the impact of each risk factor on the cost overrun prediction.

Table 4. Multiple linear regression analysis

Model	Unstandardised Coefficients		Standardised Coefficients	*T*	Sig.	95.0% Confidence Interval (CI) of β		Collinearity Statistics	
	B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
(Constant)	0.551	0.203		2.717	0.008	0.147	0.955		
(X_1) Increases in the cost of construction materials and labour	0.011	0.070	0.044	0.158	0.875	–0.129	0.151	0.164	6.068
(X_2) Change in input resources prices / variation in raw material prices	–0.029	0.112	–0.174	–0.259	0.796	–0.253	0.195	0.028	35.647
(X_3) Improper feasibility study	0.026	0.112	0.164	0.235	0.815	–0.196	0.249	0.026	38.440
(X_4) Construction cash flow problems/project funding challenges	0.032	0.067	–0.126	–0.584	0.635	–0.165	0.101	0.184	5.449
(X_5) Unexpected location/ground conditions	–0.035	0.060	–0.153	–0.584	0.561	–0.154	0.084	0.185	5.391

Note: *F*-statistic = 0.953, sig. value = 0.452, degree-of-freedom (5 and 74); R = 0.246; R^2 = 0.060 (6.0%).

Multiple linear regression model validation

The model validation process entails developing the model with the test datasets comprising 19 projects extracted from the 97 projects after removing outliers using Equation 4. The multiple linear regression equation indicated that the predicted values were not supposed to deviate from the expected percentage by more or less than 27% (as shown in Table 5). Also, the R^2 of the trained model was 0.061, which indicates that it has negligible predictive ability. This affirms the limitations of multiple linear regression models in predicting patterns in non-linear relationships among dependent and independent variables.

Table 5. Multiple linear regression training and validation results

Model	Partition	R^2	MSE	MAPE	RMSE
Cost overrun (impact)	Training	0.0605 (6.05%)	0.0408	15.83	0.2019
	Validate	0.0027 (0.27%)	0.0383	17.11	0.1957

Artificial neural network

Artificial neural network model development consists of the following steps: data gathering for tasks with regards to network creation, configuration of the network, initialisation of weight and bias, training of the network, validation of the network and analysis of data. Equation 4 provided predictive cost overruns (*Y*) based on the influence of various cost risks. The artificial neural networks model was constructed using the neural net function and incorporated independent variables: CR7, CR9, CR5, CR19, CR3 and CR13. Equation 4 featured the hidden layers, with five neurons in the first layer and three neurons in the second layer. The error function, 'sse' (sum of squared errors), with a threshold of 0.05 set for desired accuracy. Linear output was enforced for the model. Visualising the model using the plot function could offer insights into its performance and fit to the data.

```
#Build the Neural model = CRModel_2 < –neuralnet(Yc ~ CR7 + CR9 + CR5
+ CR19 + CR3 + CR13,data = train,hidden = c(5,3),err.fct = 'sse',threshold =
0.05,linear.output= T)plot(CRModel_2,rep = 'best')                    Eq. 4
```

Model development using the artificial neural networks technique

The artificial neural networks modelling process starts with determining the network architecture, followed by the learning process and, thereafter, testing the network. The nature of the problem, complexity and features of data, as well as the quantity of samples, are some elements that influence the choice of the network architecture (Sodikov, 2005). Since there is no precise method for defining artificial neural network architecture in just one attempt, it typically takes a number of tries and errors to arrive at a suitable artificial neural network architecture with less predictive error (as shown in Table 6). Hegazy, Fazio, and Moselhi (1994) suggest that the number of hidden nodes should be set at one-half of the total input and output nodes serve as a guide for this artificial neural network architecture selection.

Neural networks with the least error (MAPE, MSE and RMSE), when compared with others, were considered most accurate and were adopted. Based on these criteria, after several trials, a suitable network was found. Figure 1 presents a graphical representation of the sensitivity analysis shown in Table 6. It can be deduced from the graph that Model 2 (6–5–3–1) correlates closely with the pattern of the observed or test datasets.

Table 6. Artificial neural networks model's sensitivity test

Model	Partition	R^2	MSE	MAPE	RMSE
Model 1 (6–6–3–1)	Training	0.0499 (5.0%)	0.0415	15.90	0.2036
	Validate	0.0256 (2.6%)	0.0367	16.29	0.1915
Model 2 (6–5–3–1)	Training	0.0314 (3.1%)	0.0420	15.94	0.2051
	Validate	0.0347 (3.4%)	0.0356	15.78	0.1888
Model 3 (6–5–2–1)	Training	0.0438 (4.4%)	0.0416	16.01	0.2040
	Validate	0.0341 (3.4%)	0.0368	16.16	0.1918
Model 4 (6–4–2–1)	Training	0.0528 (5.3%)	0.0411	15.96	0.2028
	Validate	0.0012 (0.1%)	0.0367	16.39	0.1915

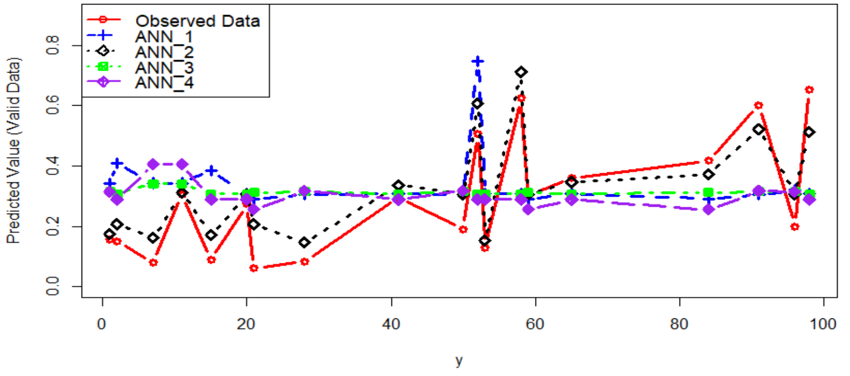

Figure 1. Artificial neural networks sensitivity analysis chart

The most suitable network architecture consisted of two hidden layers of eight nodes (6–5–3–1), as presented in Table 6 and Figure 2. The artificial neural network estimation included six inputs and one output function given by Yc, where cost risk factors were the independent variable and Yc was the dependent variable.

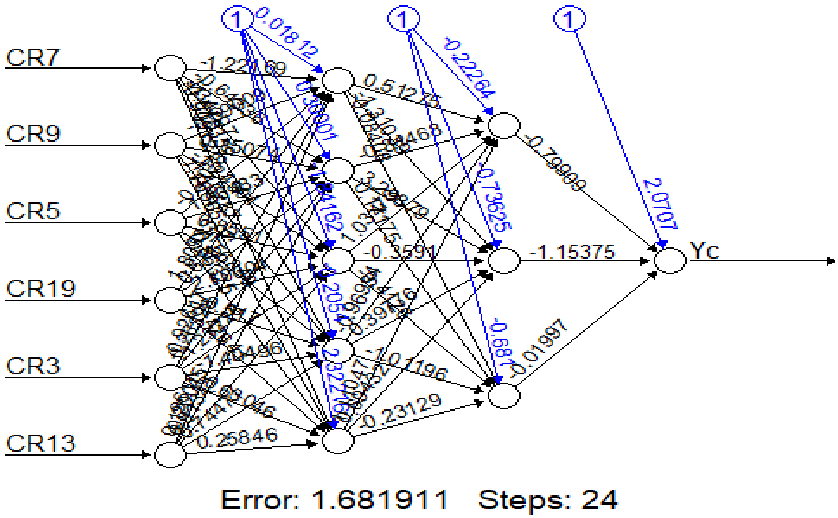

Figure 2. Artificial neural networks model architecture (6–5–3–1)

Artificial neural networks model validation

The artificial neural network models were validated by testing the results of the trained dataset against a new dataset (test dataset), and the performance was documented. One of the most crucial stages of model building is validating and verifying the model's accuracy (Dysert, 2001). A new dataset (the test dataset) was used to validate the model. It excluded the training datasets. Immediately after the model's stability was confirmed and the generated output established, the output was handled as a model that could subsequently be developed and abstracted for future use. The dataset for the model validation consisted of 19 projects extracted from the 97 projects. The predicted cost overrun was computed and compared with the cost overruns of the 19 test datasets, as presented in Table 7.

RESULT AND DISCUSSION

Highway Projects Cost-Risk Factors

The Pareto principle, also known as the 80/20 rule, dictates that 80% of effects stem from 20% of causes, guiding construction project risk management to prioritise resources towards critical risk factors. This study aligned with this principle, identifying the top six risk factors of the ranked mean scores, which collectively constituted approximately 80% of the most significant risks. These factors, such as inflation, foreign exchange fluctuations and changes in input prices, strongly impacted project costs and demanded increased attention during planning and execution. The diverse distribution of risk factors across varying risk levels underscores the multifaceted challenges in construction projects, with immediate high-level risks like inflation and exchange rate fluctuations alongside moderate or low risks like scope vagueness and tax changes. By ranking the RII scores of the time-risk factors and using the Pareto 80/20 rule, stakeholders could strategically allocate resources, enhance cost predictability, optimise project outcomes and mitigate financial losses. This emphasised the critical role of proactive risk management strategies in ensuring the success and sustainability of Nigeria's highway construction projects.

Multiple Linear Regression Model Development and Validation

The results of the multiple linear regression model development, presented in Table 4, highlighted the presence of a significant correlation between variables, leading to the elimination of one predictor. Additionally, the analysis revealed high multicollinearity among predictors, indicated by elevated VIF values and low tolerance values. Table 7 reveals a notable discrepancy between predicted and actual values, with a substantial MAPE of 17.11%. Despite these limitations, the result was considered satisfactory, given the intricate nature of construction project management and the inherent uncertainties involved.

Artificial Neural Networks Model Development and Validation

The validation of the artificial neural networks cost model revealed insights into its predictive accuracy. Across training and validation phases, the model showed a consistent MAPE and RMSE value around 15.78% and 0.1888, respectively (as shown in Table 6). The artificial neural networks model's ability to generalise to unseen data suggested its practical use in real life.

Multiple Linear Regression and Artificial Neural Networks Model Comparison

After the removal of outliers from the first 103 historical datasets, the multiple linear regression and artificial neural networks prediction models were developed from 97 datasets. The datasets were split into 80:20 training and test datasets. From the results presented in Table 6, the preferred artificial neural networks model (artificial neural networks architecture of 6–5–3–1) had a MAPE of 15.78%. The result indicated that the model had good predictive accuracy, as suggested by Lewis (1982), which states that the MAPE of a good model should not exceed 20%. Because the artificial neural networks model can predict the future, it is possible to get very accurate estimates of how much highway projects will cost in the future. The test dataset was then utilised to validate the model. Although the multiple linear regression model had a MAPE of 0.1711, the five chosen predictors (i.e., $R^2 = 0.0605$) only accounted for 6.05% of the difference in cost overrun. Table 7 presented a comparison between the test dataset, which was the actual projects, cost, or cost overrun and the predicted data of the multiple linear regression model and the artificial neural networks model, respectively. For the artificial neural networks model, it was found that 58% of the test data had a percentage error of less than 20%, while the multiple linear regression model had 21% of the test data with a percentage error of less than 20%. Furthermore, the results in Table 7 are represented graphically in Figure 3 for the purpose of visual comparison. The parentage error between the actual overrun (test data) and the predicted overrun was computed using Equation 5.

$$P_e = \frac{(APC - PPC)}{APC} \times 100 \quad \text{Eq. 5}$$

where P_e = Percentage error, APC = Actual project cost and PPC = Predicted project cost.

Table 7. The summary of actual cost overrun, multiple linear regression (MLR) predictive cost overrun and artificial neural networks (ANN) predictive cost overrun

Test Data S/N (Refer Appendix)	Actual Project Cost (Cost Overrun Data)	Multiple Linear Regression		Artificial Neural Networks	
		Predicted	Percentage Error (%)	Predicted	Percentage Error (%)
1	0.1579	0.316746	–50.15	0.173874	–9.19
2	0.1511	0.325072	–53.52	0.207126	–27.05
7	0.0810	0.404733	–79.99	0.162570	–50.18
11	0.3063	0.404733	–24.32	0.312257	–1.91
15	0.0897	0.325072	–72.41	0.171264	–47.62
20	0.2744	0.325072	–15.59	0.307126	–10.66
21	0.0609	0.252623	–75.89	0.207232	–70.61
30	0.0845	0.293227	–71.18	0.146929	–42.49
39	0.2950	0.325072	–9.25	0.337126	–12.50
48	0.1917	0.293227	–34.62	0.306929	–37.54
50	0.5070	0.325072	55.97	0.607126	–16.49
51	0.1302	0.325072	–59.95	0.153713	–15.30
56	0.6245	0.325072	92.11	0.712643	–12.37
57	0.3021	0.252623	19.59	0.307232	–1.67
63	0.3599	0.325072	10.71	0.347126	3.68
82	0.4172	0.252623	65.15	0.37232	12.05
89	0.6016	0.293227	105.17	0.522854	15.06
94	0.2003	0.316746	–36.76	0.307387	–34.84
96	0.6533	0.325072	100.97	0.512643	27.44

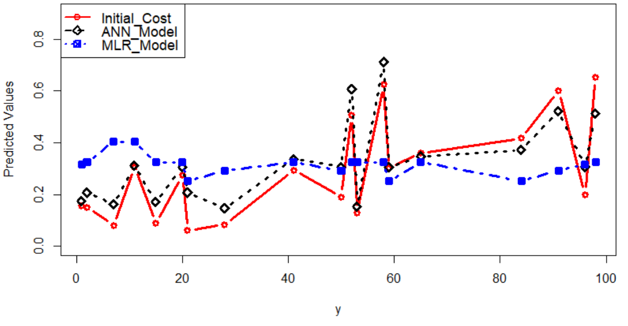

Figure 3. Initial cost overrun, multiple linear regression predicted cost overruns and artificial neural networks predicted cost overruns chart

Models' Application

Figure 4 depicts the application of the artificial neural networks cost predictive model to estimate the costs for new highway projects, beginning with defining goals and progressing to project design and data processing. The activation of the pre-developed artificial neural networks model was done after choosing the critical cost prediction parameters, such as cost and risk. After activation, the model used the projected final cost of the new project to determine new project costs. Regular assessments and adjustments were necessary to accommodate any changes or new data that might influence the project's cost. The project finished in knowledge transfer, which could guide future efforts and promote organisational learning. The project cycle ended after the complete integration of the model and dissemination of the resulting insights. Figure 4 provides a flowchart that highlights a structured and data-driven approach to cost estimation, emphasising continuous improvement and knowledge sharing.

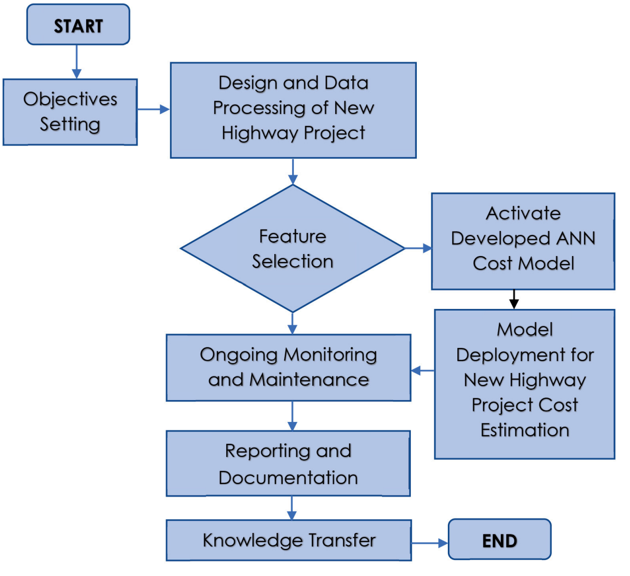

Figure 4. Artificial neural network's cost predictive model application flowchart

CONCLUSIONS

In this study, critical cost-risks of highway projects were identified and predictive cost estimating models were developed to determine the impact of critical cost-risks on the actual costs of highway projects at an early phase of the project.

Summary of Findings

Major cost-risk factors impact highway projects by increasing the cost of construction materials and labour, depending on the variations in a country's exchange rate and input resource prices. In Nigeria's federal highway projects, their cost estimation is highly prone to inflation and responds to the volatility of the country's currency and foreign exchange rate.

This study found that the artificial neural networks model performed much better than the multiple linear regression model, with an accuracy of 84.22%. The result further clarifies the superiority of the artificial neural networks model over the multiple linear regression. Researchers could be guided by the findings to explore the application of artificial neural networks further in similar studies.

This study also presented a valuable tool for consultants and project managers to appreciate how different risk configurations impact the cost of highway projects.

The results of this study, which revealed an increase in the cost of construction materials and labour, foreign exchange rate variation and changes in input resource prices, are important findings that can encourage policymakers to promptly and adequately release certified fees and valuations to enhance cash flow during the project construction life cycle.

Significance of Study

The study's significance lies in its potential to revolutionise project management in the construction industry, particularly in Nigerian highway projects. By introducing a predictive model that utilises artificial neural networks to forecast the impact of risk on project costs, the study offers a proactive solution to a longstanding problem. This approach enables highway contractors and project managers to identify critical cost-risk factors early in the planning phase, facilitating the adoption of efficient risk management strategies. The study also suggests using data-driven methods instead of guesswork for traditional estimation methods. This will lead to more accurate project estimates and lessen the impact of cost overruns.

Contributions

The developed model can enhance decision-making in project management to aid highway contractors and project managers in identifying critical cost-risk factors for Nigerian highway projects. This enables the adoption of an efficient risk management strategy. Additionally, the study presented a proactive, data-driven approach to project cost determination, contrasting with traditional estimation methods that are reliant on assumptions for contingency planning. Implementing the proposed approach from this study promises to significantly improve cost performance on Nigerian highway projects.

Recommendations

Professionals and experts in the construction industry should advocate and embrace the use of artificial neural networks to forecast the influence of risk on the actual cost of highway projects. Implementing this approach during the conceptual and planning stages of highway projects will facilitate more realistic project estimates and reduce the occurrence and impact of cost overruns in Nigerian highway projects.

ACKNOWLEDGEMENTS

The authors acknowledge Universiti Teknologi Malaysia for facilitating the research environment. Also, we are grateful to the respondents for furnishing the essential data for this study. This work was supported by the Ministry of Higher Education's Fundamental Research Grant Scheme (grant number R.J130000.7852.5F025).

REFERENCES

Al-Suhaili, R., Saco, Z.M. and Al-Zwainy, M. (2010). Using artificial neural networks for predicting the total cost of highway projects in Iraq. Paper presented at the 2nd Regional Conference for Engineering Science. College of Engineering. Al-Nahrain University, Iraq, 1–2 December.

Al-Zwainy, F.S. and Aidan, F.M.S. (2017). Forecasting the cost of structure of infrastructure projects utilising artificial neural network model (highway projects as case study). *Indian Journal of Science and Technology*, 10(20): 1–12. https://doi.org/10.17485/ijst/2017/v10i20/108567

Anigbogu, N.A., Ahmad, Z.B. and Molwus, J.J. (2019). Cost overruns on federal capital territory authority road construction projects. *Journal of the Environment*, 13(1): 1–14.

Awuku, B., Asa, E., Baffoe-Twum, E. and Essegbey, A. (2022). Conceptual cost estimation of highway bid items: A systematic literature review. *Engineering, Construction and Architectural Management*, 31(3): 1187–1221. https:/doi.org/10.1108/ECAM-03-2022-0266

Barros, L.B., Marcy, M. and Carvalho, M.T.M. (2018). Construction cost estimation of Brazilian highways using artificial neural networks. *International Journal of Structural and Civil Engineering Research*, 7(3): 283–289. https://doi.org/10.18178/ijscer.7.3.283-289

Changali, S., Mohammad, A. and Van Nieuwland, M. (2015). The construction productivity imperative. Available at: https://www.mckinsey.com/capabilities/operations/our-insights/the-construction-productivity-imperative [Accessed on 23 July 2024].

Dysert, L.R. (2001). Developing a parametric model for estimating process control costs. *International Journal of Construction Engineering*, 43(2): 31–34.

Elbeltagi, H., Hosny, O., Abdel-Razek, R. and El-Fitory, A. (2014). Conceptual cost estimate of Libyan highway projects using artificial neural networks. *Journal of Engineering Research and Application*, 4(8): 56–66.

El-Kholy, A.M. (2019). Exploring the best ANN model based on four paradigms to predict delay and cost overrun percentages of highway projects. *International Journal of Construction Management*, 21(7): 694–712. https://doi.org/10.1080/15623599.2019.1580001

El-Sayegh, S.M. and Mansour, M.H. (2015). Risk assessment and allocation in highway construction projects in the UAE. *Journal of Management in Engineering*, 31(6): 04015004. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000365

Federal Ministry of Power, Works and Housing (2017). List of on-going federal highway projects. Available at: https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://worksandhousing.gov.ng/themes/front_end_themes_01/images/uploads_images/1562351340.pdf&ved=2ahUKEwiNsMzC-LyHAxUdZ0EAHXx8AP8QFnoECBkQAQ&usg=AOvVaw0K7XRMEuv8IjsVOPwxEszX [Accessed 15 December 2021]

Fernando, N., Dishan, T.A.K. and Zhang, H. (2023). An artificial neural network (ANN) approach for early cost estimation of concrete bridge systems in developing countries: The case of Sri Lanka. *Journal of Financial Management of Property and Construction*, 29(1): 23–51. https://doi.org/10.1108/JFMPC-09-2022-0048

Glymis, E., Kanelakis, A., Aretoulis, G.N. and Mastoras, T. (2017). Predicting highway projects' actual duration using neural networks. Paper presented at the Lean and Computing in Construction Congress: Joint Conference in Computing Construction (LC3-2017). Heraklion, Greece, 4–7 July. https://doi.org/10.24928/JC3-2017/0260

Gondia, A., Siam, A., El-Dakhakhni, W. and Nassar, A.H. (2019). Machine learning algorithms for construction projects delay risk prediction. *Journal of Construction Engineering and Management*, 146(1): 04019085. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001736

Grosfeld-Nir, A., Ronen, B. and Kozlovsky, N. (2007). The Pareto managerial principle: When does it apply? *International Journal of Production Research*, 45(10): 2317–2325. https://doi.org/10.1080/00207540600818203

Hamma-Adama, M., Ogunsemi, M., Mashwama, N.X., Ahmad, A.B.S. and Abba, H.A. (2021). Assessment of risks associated with road infrastructure development in developing countries. *Global Scientific Journal*, 9(7): 1223–1229.

Hegazy, T., Fazio, P. and Moselhi, O. (1994). Developing practical neural network applications using back-propagation. *Microcomputers in Civil Engineering*, 9(2): 145–159. https://doi.org/10.1111/j.1467-8667.1994.tb00369.x

Ikechukwu, A.N. and Akiohnbare, W. (2017). Factors affecting cost of construction in Nigeria. *International Journal of Advancement in Engineering Technology, Management and Applied Science*, 4(6): 48–61

Kovacevic, M., Ivanisevic, N., Petronijevic, P. and Despotovic, V. (2021). Construction cost estimation of reinforced and pre-stressed concrete bridges using machine learning. *Građevinar*, 73(1): 1–13. https://doi.org/10.14256/JCE.2738.2019

Leo-Olagbaye, F. and Odeyinka, H.A. (2020). An assessment of risk impact on road projects in Osun State, Nigeria. *Built Environment Project and Asset Management*, 10(5): 673–691. https://doi.org/10.1108/BEPAM-04-2019-0040

Lewis, C.D. (1982). *International and Business Forecasting Methods*. London: Butterworths. https://doi.org/10.1002/FOR.3980020210

Love, P., Edwards, D. and Iran, Z. (2012). Moving beyond optimism bias and strategic misrepresentation: An explanation for social infrastructure project cost overruns. *IEEE Transportation Engineering Management*, 59(4): 590–571. https://doi.org/10.1109/TEM.2011.2163628

Okate, A. and Kakade, V. (2019). Risk management in road construction projects. High-volume roads. In *Proceedings of Sustainable Infrastructure Development and Management (SIDM)*. Maharashtra, India: Visvesvaraya National Institute of Technology. https://doi.org/10.2139/ssm.3375904

Pineda-Jaramillo, J.D., Insa, R. and Martínez, P. (2017). Modelling the energy consumption of trains applying neural networks. *Journal of Rail and Rapid Transit*, 232(3): 816–823. https://doi.org/10.1177/0954409717694522

Rizwan, M. and Iqbal, M. (2011). Application of 80/20 rule in software engineering rapid application development (RAD) model. In J.M. Zain, W.M.W. Mohd, and E. El-Qawasmeh (eds.), *Software Engineering and Computer Systems. ICSECS 2011. Communications in Computer and Information Science, Vol. 181*. Berlin: Springer, 518–532. https://doi.org/10.1007/978-3-642-22203-0_45

Sarker, I.H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, 2(3): 1–21. https://doi.org/10.1007/s42979-021-00592-x
Sodikov, J. (2005). Cost estimation of highway projects in developing countries: Artificial neural network approach. *Journal of the Eastern Asia Society for Transportation Studies*, 6: 1036–1047. https://doi.org/10.11175/easts.6.1036
Tariq, J. and Gardezi, S.S. (2023). Study the delays and conflicts for construction projects and their mutual relationship: A review. *Ain Shams Engineering Journal*, 14(1): 101815. https://doi.org/10.1016/j.asej.2022.101815
Thaseena, T. and Vishnu, K. (2017). Analysis of cost overruns in highway construction projects using multiple regression and artificial neural networks. *International Research Journal of Engineering and Technology*, 4(4): 2332–2336.
Tijanić, K., Car-Pušić, D. and Šperac, M. (2019). Cost estimation in road construction using artificial neural networks. *Neural Computing and Applications*, 32(13): 9343–9355. https://doi.org/10.1007/s00521-019-04443-y

APPENDIX

Highway project historical/Experiment data

No.	Length (km)	Initial Contract Sum (NGN Million)	Actual Contract Sum (NGN Million)	Cost Overruns (%)	Adjustment for Data Compatibility (%)
1	19.20	3,200	3,705	15.79	0.1579
2	31.00	2,500	2,878	15.11	0.1511
3	22.00	3,502	4,604	31.48	0.3148
4	10.50	14,990	17,490	16.68	0.1668
5	5.60	6,210	6,338	2.05	0.0205
6	7.20	8,661	10,025	15.75	0.1575
7	84.00	96,304	104,110	8.10	0.0810
8	72.70	3,993	5,111	28.00	0.2800
9	166.02	65,223	72,619	11.34	0.1134
10	24.00	1,535	2,132	38.87	0.3887
11	27.60	2,693	3,518	30.63	0.3063
12	16.90	188	323	71.57	0.7157
13	52.00	47,504	73,525	54.78	0.5478
14	30.00	999	1,352	35.36	0.3536
15	75.00	3,536	3,853	8.97	0.0897
16	5.20	873	1,229	40.73	0.4073
17	52.00	47,504	52,141	9.76	0.0976
18	32.20	2,600	2,786	7.14	0.0714
19	46.00	2,137	3,169	48.30	0.4830
20	0.82	250	319	27.44	0.2744
21	51.00	37,500	39,785	6.09	0.0609
22	21.00	2,995	4,019	34.19	0.3419
23	33.49	9,998	12,473	24.75	0.2475

No.	Length (km)	Initial Contract Sum (NGN Million)	Actual Contract Sum (NGN Million)	Cost Overruns (%)	Adjustment for Data Compatibility (%)
24	25.00	4,613	6,431	39.41	0.3941
25	30.00	5,209	7,448	42.99	0.4299
26	25.50	5,245	6,780	29.27	0.2927
27	337.00	64,125	91,590	42.83	0.4283
28	83.01	65,220	74,876	14.81	0.1481
29	0.50	139	144	3.47	0.0347
30	55.43	8,720	9,504	9.00	0.0900
31	338.47	29,922	41,562	38.90	0.3890
32	470.32	44,884	66,000	47.05	0.4705
33	18.70	113	130	14.88	0.1488
34	46.00	2,093	3,123	49.22	0.4922
35	59.50	39,550	43,973	11.18	0.1118
36	7.00	515	542	5.26	0.0526
37	10.00	991	1,250	26.16	0.2616
38	49.00	5,092	6,657	30.73	0.3073
39	49.00	4,614	5,975	29.50	0.2950
40	10.50	600	750	25.00	0.2500
41	17.00	1,277	1,683	31.78	0.3178
42	10.00	251	416	65.84	0.6584
43	13.50	1,787	2,410	34.90	0.3490
44	75.00	8,965	9,623	7.34	0.0734
45	40.27	11,603	12,519	7.90	0.0790
46	39.00	3,287	4,325	31.59	0.3159
47	58.00	200	384	91.81	0.9181
48	22.00	4,208	5,014	19.17	0.1917
49	76.00	4,207	5,163	22.73	0.2273
50	25.80	3,296	4,967	50.70	0.5070
51	19.50	66,830	75,529	13.02	0.1302
52	42.00	11,228	18,058	60.83	0.6083
53	9.00	982	1,305	32.83	0.3283
54	50.00	11,987	19,352	61.44	0.6144
55	58.59	11,664	18,877	61.84	0.6184
56	49.36	9,697	15,753	62.45	0.6245
57	93.60	14,587	18,994	30.21	0.3021
58	5.00	1,320	1,849	40.04	0.4004
59	88.00	7,935	9,518	19.95	0.1995
60	98.00	4,424	6,254	41.36	0.4136
61	13.30	3,383	4,571	35.13	0.3513
62	82.00	5,402	8,883	64.44	0.6444
63	5.50	1,185	1,611	35.99	0.3599
64	100.00	8,969	13,827	54.16	0.5416

No.	Length (km)	Initial Contract Sum (NGN Million)	Actual Contract Sum (NGN Million)	Cost Overruns (%)	Adjustment for Data Compatibility (%)
65	53.40	8,847	9,672	9.33	0.0933
66	40.00	9,951	12,286	23.46	0.2346
67	24.50	3,698	3,977	7.54	0.0754
68	24.30	2,253	4,394	95.02	0.9502
69	15.00	3,076	3,477	13.01	0.1301
70	117.78	35,841	52,015	45.13	0.4513
71	101.84	29,100	39,458	35.59	0.3559
72	38.20	7,130	8,820	23.70	0.2370
73	145.11	39,999	47,869	19.67	0.1967
74	73.00	5,020	9,998	99.17	0.9917
75	20.00	1,240	1,592	28.32	0.2832
76	122.00	5,720	7,764	35.74	0.3574
77	36.28	9,881	12,384	25.34	0.2534
78	68.00	7,257	8,135	12.11	0.1211
79	69.00	5,156	5,808	12.64	0.1264
80	6.50	984	1,176	19.46	0.1946
81	25.00	6,582	7,507	14.05	0.1405
82	96.24	30,250	42,870	41.72	0.4172
83	25.00	6,582	7,505	14.02	0.1402
84	30.00	2,319	2,914	25.65	0.2565
85	32.80	7,953	9,229	16.04	0.1604
86	14.08	610	688	12.85	0.1285
87	13.59	1,764	2,468	39.90	0.3990
88	65.00	3,792	5,037	32.82	0.3282
89	64.90	16,000	25,625	60.16	0.6016
90	10.20	989	1,020	3.15	0.0315
91	21.30	4,394	5,420	23.36	0.2336
92	26.60	13,227	18,772	41.92	0.4192
93	105.00	37,047	49,195	32.79	0.3279
94	104.00	7,942	9,533	20.03	0.2003
95	210.00	3,000	4,397	46.58	0.4658
96	296.00	10,560	17,459	65.33	0.6533
97	11.40	2,256	3,548	57.25	0.5725