NRC Research Press

ARTICLE

Modelling a cost profile for road projects

Opeoluwa Akinradewo, Clinton Aigbavboa, Ayodeji Oke, and Harrison Coffie

Abstract: One of the vital success elements of a construction project is the accuracy of the estimation of construction cost. This study is aimed at developing a cost profile for road projects in Ghana. Pro forma was designed to retrieve historical cost data of completed road projects in Ghana. The pro forma retrieved data such as the initial budgeted cost and final construction cost of road projects, location of road projects, features of road projects, the scope of road projects (new project, renovation work, upgrade work or replacement work), type of road projects, and classification of road projects. Cost data were analyzed using descriptive analysis and probability distributions such as cumulative density functions and probability density functions. From the findings, estimates prepared for road projects in Ghana can be expected to be below the final construction cost by approximately 20% as most of the completed road projects in Ghana experience cost overrun.

Key words: cost performance, cost profiling, Ghana, probability distribution, road infrastructure.

Résumé : L'un des éléments essentiels du succès d'un projet de construction est la justesse de l'estimation des coûts de construction. Cette étude vise à établir un profil des coûts pour les projets routiers au Ghana. « pro forma » a été conçu pour récupérer des données historiques sur les coûts des projets routiers achevés au Ghana. Les données pro forma récupérées telles que le coût budgété initial et le coût final de construction des projets routiers, l'emplacement, les caractéristiques et la portée des projets routiers (nouveaux projets, travaux de rénovation, travaux de modernisation ou de remplacement), type et classification des projets routiers. Les données sur les coûts ont été examinées à l'aide d'une analyse descriptive et de distributions de probabilités comme les fonctions de densité cumulative et les fonctions de densité de probabilité. D'après les résultats, on peut s'attendre à ce que les estimations préparées pour les projets routiers au Ghana soient inférieures d'environ 20 % au coût de construction final, étant donné que la plupart des projets routiers achevés au Ghana connaissent des dépassements de coûts. [Traduit par la Rédaction]

Mots-clés : évolution des coûts, profil des coûts, Ghana, distribution des probabilités, infrastructure routière.

Introduction

It is a general phenomenon that the construction sector plays a key part in the development of the socioeconomic growth of a country. These national socioeconomic development goals include making necessary infrastructure (highways, schools, hospitals, houses, townships, roads, railways, airports, seaports, power systems, agriculture systems, telecommunications, among others) available, providing employment opportunities and also shelter for the society (Chileshe and Berko 2010; Osei 2013). A large part of the economy of a country is attributed to the construction industry, which contributes up to 10% of the gross domestic product (GDP) and also employs approximately 10% of the working population. The construction industry deals with the creation, repair, renovation or extension of infrastructure, land improvements, buildings of engineering nature referred to as fixed assets of the country. "Close to half of the gross fixed capital formation of an economy is also constituted by the construction industry which shows the relevance of the construction industry is to the country's economy" (Ofori-Kuragu et al. 2016).

Kim et al. (2004) posited that one of the vital success elements of a construction project is the accuracy of the estimation of construction cost. Even though estimation for construction projects is of great importance, it is neither straightforward nor a simple task. Preparation of preliminary estimate is tedious due to the lack of full project details in the early phases of the construction project (Hegazy 2002). Chou et al. (2005) also supported the claims putting forward that determining the viability of a project even before its initiation depends largely on the preliminary estimates. This is because a significant loss can be made by a contractor or project abandonment by the client in the case of cost underestimation. Liu and Zhu (2007) in the same vein posited that the capability to accurately predict a project delivery cost is central to the competition when a bid is submitted for construction projects. According to Swei et al. (2017), a review of the trends in construction cost estimation shows a low level of accuracy of preliminary estimates compared to final construction cost in the last five decades. Countless cost estimation models have therefore been developed for determining preliminary estimates for construction projects. However, only Ogungbile et al. (2018) developed a model for an African country (Nigeria) with the focus of the cost model being the parametric estimating technique. The model focused on the following activities of road infrastructure construction: asphalt work, base course, concrete drains, earthwork, stone pitching of embankments, and scarification of the existing bituminous surface. To this end, this study is imperative in achieving preliminary estimates that are more reliable for road projects funded by the government in the Ghanaian construction industry. Such that cost overrun can be ameliorated by adequately evaluating contingency costs for road

Received 19 November 2019. Accepted 20 February 2020.

O. Akinradewo, C. Aigbavboa, A. Oke, and H. Coffie. SARChl in Sustainable Construction Management and Leadership in the Built Environment, Faculty of Engineering and the Built Environment, University of Johannesburg, Johannesburg, South Africa.

Corresponding author: Opeoluwa Akinradewo (email: opeakinradewo@gmail.com).

projects before the project execution. This will be achieved by developing a cost profile for road projects in Ghana which will assist estimators in proper budgeting for road projects. This will enable the clients, sponsoring organizations, as well as the project team, to have a general overview of the cost implication of a road infrastructure to be embarked upon. This will give room for making necessary projections to ensure the infrastructure is achieved.

Cost profiling

The word profiling is used in different contexts for different purposes among which are marketing research, healthcare application, computer engineering, mathematics, forensic biometrics, actuarial justice, and supply chain management among others. There is one major connection with the different applications of profiling, which is the "use of algorithms or other techniques to create, discover or construct knowledge/patterns/correlations from huge sets of data" (Seneviratne and Levy 2010). The correlation shows the relationship between data and makes a prediction based on past events and happenings. Profiling in essence, therefore, is an inductive way of generating knowledge showing the probability that something will happen in the future the same way it did in the past (Hildebrandt and Gutwirth 2008). From the foregoing, cost profiling can be defined as the procedure of determining the correlations between data in cost databases for a set subject that can be used to identify and represent a future cost for a similar subject. According to Hildebrandt and Gutwirth (2008) understanding what profiling entails requires knowing the purpose of profiling. The purpose is to assess the opportunities and risks in the recorded data for future predictions. Profiling seeks to address the complexity and diversity of real data through categorization of the different attributes of such data (Canhoto and Backhouse 2008).

Profiling employs the technique of calculating similarities between collated data. These similarities are also referred to as correlation that exists between the set of available data. For a set of people living in a settlement, there can be a correlation found among them such as a particular disease or income level, among others. For the correlation result to be valid, data must be gathered for a given extended period of time and analyzed (Hildebrandt and Gutwirth 2008). Profiling can be distributive or non-distributive. When all the data that makes up the group profiles have the same attributes, it is said to be distributive. Application of this profile can be done without affecting any member of the group. The non-distributive profile is associated with a group of data whereby all its members do not have the same attribute (Edens 2001). According to Hildebrandt and Gutwirth (2008), to tackle the continuous growth of knowledge complexities and unpredictability of infrastructures, there is a need for the development of profiling technologies. This will assist in anticipating adequately the results of alternative courses of action. To achieve profiling for the cost of construction projects, executed construction project details are captured and stored over a given time period. The stored data are used for profiling by modelling performance and measuring deviation from the performance modelled (Canhoto and Backhouse 2008).

Researchers have employed the use of cost profiling for different purposes in the construction industry to achieve set objectives. Chou et al. (2005) developed a cost profile for preparing a "preliminary cost estimate for highway bridge replacement projects". The cost profile was developed using 546 highway projects to achieve "an alternate probabilistic cost estimating that could offer the confidence bounds and control the desired error". Love et al. (2013) carried out a study on "Determining the Probability of Project Cost Overruns" for the Australian construction industry. The study employed 276 engineering and construction projects in developing the cost profile. The cost profile model is useful in "producing realistic probabilities of cost overruns, which should be incorporated into a construction cost contingency" for the engineering and construction projects in Australia. Love et al. (2015) employed cost profiling in "Estimating Construction Contingency: Accommodating the Potential for Cost Overruns in Road Construction Projects". This was achieved by using 49 road projects executed in Australia to "provide a more robust approach for governing a construction contingency for road construction projects so that they can be procured successfully". Also, Love et al. (2018) examined 1093 water projects executed in the United Kingdom between 2002 and 2012 to develop a cost profile. The cost profile achieved is used in "providing decision-makers with a reliable basis to determine an appropriate contingency level and therefore they make a positive contribution to improving the management of risk".

Research methodology

An exploratory study was implemented to achieve the set objective of this research. Historical cost data of road projects executed between 2014 and 2019 were gathered using pro forma to explore the determinants for cost profiling of road projects and develop a cost profile for government-funded road projects in Ghana. The pro forma, which is a data form for collecting secondary data, was designed to retrieve data such as the budgeted cost of construction and final construction cost of road projects, location of road projects, features of road projects (median strip, road length, road width, type of pavement), the scope of the road projects (new project, renovation work, upgrade work or replacement work), and classification of road projects (trunk, feeder, urban). These data were collected from the Ghana Highway Authority, the Department of Urban Roads and the Department of Feeder Roads in Ghana. To ensure there is enough data to justify the model to be developed, cost data of government-funded road projects were retrieved from all the regions in Ghana. The pro forma retrieved 166 road project cost data which were found suitable for analysis as all the necessary information for the cost profile model to be developed were given. The road project costs retrieved in Ghana Cedis (₵) were converted to US Dollar ($) because it is a globally recognized monetary unit. A diagrammatic representation of how this study carried out the cost profile development is shown in Fig. 1.

Historical cost data were analyzed using descriptive statistics such as mean (M) which is used to determine the average cost by dividing the cost deviation sum by the total count; standard deviation (SD) which is used to observe the deviation in the cost data obtained from the average cost by calculating the variance of each value which is then squared and totalled before it is divided by the data count; kurtosis which is used to measure the amount of probability in the cost data; skewness which is used to measure the symmetry in the cost data distribution; and interquartile range which is used to measure the spread of the cost data from the central values. These descriptive statistics were carried out to ensure the normalization of the retrieved data before they are used for the cost profile modelling. Cost underrun or overrun on the executed road projects were determined by finding the difference between the actual construction cost and the initial estimate (award contract sum). To formulate the model, a probability function was employed. Probability can be classified into non-parametric and parametric distributions. When it considers a single continuous variable at a time, it is referred to as univariate distribution while when it considers two or more variables at a time, it is referred to as multi-variate distribution (Rossi and Deutsch 2014). Cumulative density function (CDF) is a general way of conveying the limited knowledge of a single continuous variable which is expressed mathematically with interval probability of X occurring in an interval from a to b (where $b > a$) as shown in eq. 1. Probability density function (PDF) on the other hand is derived from CDF by applying differentiation mathematical

Fig. 1. Cost profile development process.

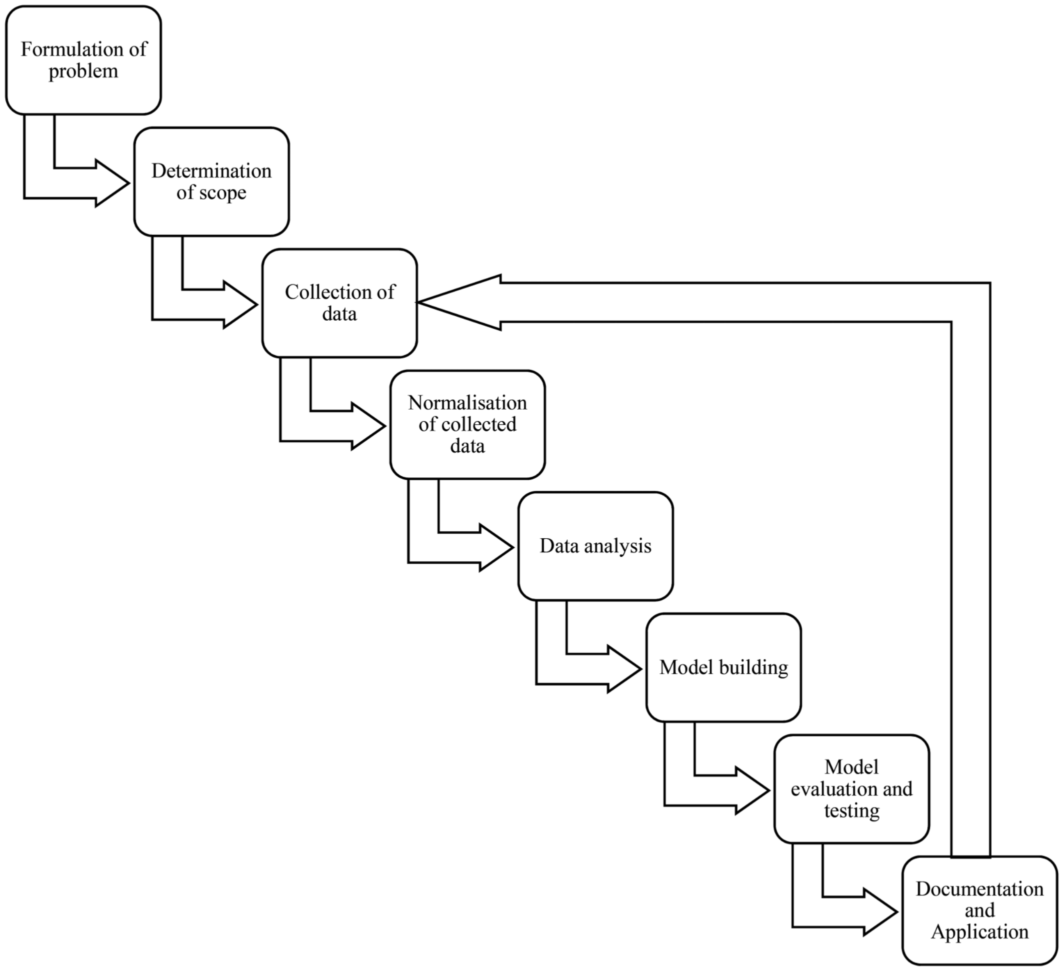


Table 1. Road project characteristics.

	Variables	Frequency	Percent
Location	North	27	16.3
	South	53	31.9
	East	42	25.3
	West	44	26.5
	Total	**166**	**100.0**
Median strip	Yes	50	30.1
	No	116	69.9
	Total	**166**	**100.0**
Pavement type	Paved road	151	91.0
	Unpaved road	15	9.0
	Total	**166**	**100.0**
Scope	New	11	6.6
	Renovation	84	50.6
	Upgrade	58	34.9
	Replacement	13	7.8
	Total	**166**	**100.0**
Road classification	Trunk	28	16.9
	Feeder	114	68.7
	Urban	24	14.5
	Total	**166**	**100.0**

Table 2. Descriptive statistics for project cost details.

Statistics (%)	Cost performance (%) (N = 166)	Cost overrun (%) (N = 122)	Cost underrun (%) (N = 36)
Mean	11.1373	17.0709	−6.4958
Std. deviation	22.68710	23.67456	4.23106
Variance	514.705	560.485	17.902
Range	144.47	129.05	15.24
Interquartile range	12.68	10.83	7.19
95% confidence interval for mean			
Lower bound	7.6606	12.8275	−7.9274
Upper bound	14.6141	21.3143	−5.0642
Minimum	−15.41	0.01	−15.41
Median	5.4250	8.8250	−7.2400
Maximum	129.06	129.06	−0.17
Kurtosis	7.671	6.282	−0.737
Skewness	2.608	2.516	−0.069

theorem expressed mathematically with interval probability of X occurring in an interval from a to b (where $b > a$) as shown in eq. 2. A cumulative histogram is useful in showing all the data of a CDF and PDF in one single plot while a probability plot will show the graph of a CDF and PDF. The probability plot shows the changes in the slope of the data thereby assisting in interpretation

Table 3. Relationship between road project characteristics.

	CP	CO	CU	L	RL	RW	MS	Pt	S	EC
CP	1									
CO	1.000*	1								
CU	1.000*	.‡	1							
L	–0.134	–0.129	0.064	1						
RL	–0.096	–0.083	–0.130	−0.272*	1					
RW	0.157†	0.113	–0.035	–0.077	0.005	1				
MS	–0.041	0.038	0.158	0.164†	–0.049	−0.554*	1			
PT	–0.070	–0.071	0.063	0.034	−0.288*	–0.081	0.115	1		
S	0.072	0.178†	0.186	–0.144	–0.046	0.048	–0.018	–0.046	1	
EC	–0.069	–0.151	0.469*	–0.078	0.219*	−0.206*	0.207*	–0.062	0.011	1

Note: CP = cost performance; CO = cost overrun; CU = cost underrun; L = location; RL = road length; RW = road width; MS = median strip; PT = pavement type; S = scope; RC = road classification.

*Correlation is significant at the 0.01 level (2-tailed).

†Correlation is significant at the 0.05 level (2-tailed).

‡Cannot be computed because at least one of the variables is constant.

and also used in checking the distribution model for normal distribution or lognormal distribution (Rossi and Deutsch 2014). According to Love et al. (2013), CDF and PDF are applicable to cost overrun determination on construction projects, which has been used by several authors for this purpose (for example: Flyvbjerg 2007; Love et al. 2013, 2015, 2016, 2018). The choice of CDF and PDF for this purpose as submitted by Love et al. (2013) relies solely on the shortcoming of a normal distribution in accurately modelling a right or left-skewed data such as a cost data. This is because normal distribution is symmetric concerning its mean value. Love et al. (2013) further stressed that “a heavy-tailed distribution model such as Cauchy” is more accurate in describing a cost data even if it is symmetric in nature. CDF and PDF are calculated therefore, using the following eqs. 1 and 2 respectively.

$$\int_a^b f(x)\mathrm{d}x = P(a \le X \le b) \tag{1}$$

$$F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t \tag{2}$$

Pearson’s correlation coefficient was used to explore the observed differences between the cost performance, cost overrun, cost underrun, project location, road length, width, features, classification, and scope. This is to determine the relationship that exists between the continuous and quantitative variables. One-way analysis of variance (ANOVA) was used to examine the differences between the road project characteristics used for the cost profile while a Tukey’s honest significant difference (HSD) post-hoc test was used to identify where the observed differences may have existed within the samples. Understanding the relationship as well as the differences between the road project characteristics will give an insight into what characteristic plays a major role in the cost profile to be developed. As stated earlier, cumulative density functions (CDF) and probability density functions (PDF) were therefore employed for the probability distribution of cost performance (the difference between the budgeted cost and the final construction cost) using cumulative histogram and probability plots. This is because CDF and PDF will reveal the probability of encountering cost overrun or underrun for future road projects in Ghana.

Findings and discussion

The cost data retrieved indicated the characteristics of the road projects such as budgeted construction cost and final construction cost, scheduled completion time and final completion time, location, features (median strip, road width, type of pavement), project scope (new project, renovation work, upgrade work or replacement work), and road classification (trunk, feeder, urban). Table 1 shows the frequency and percentage of each characteristic.

Table 4. One-way between the group ANOVA.

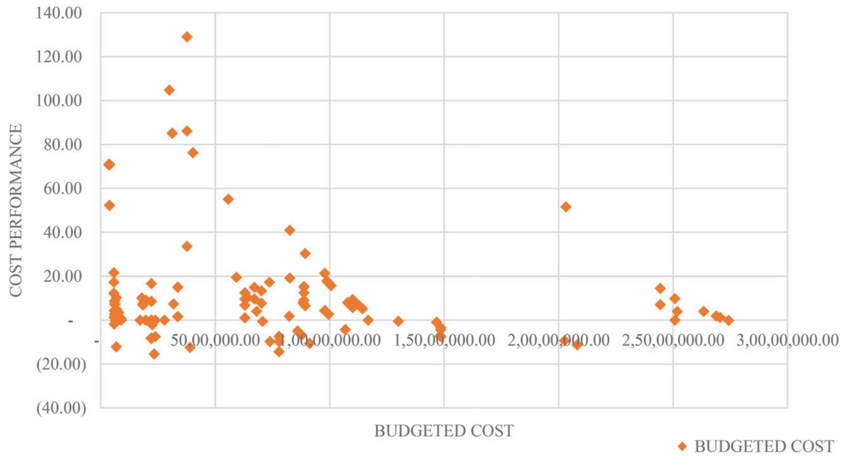
	Sum of squares	Mean square	F	Sig.
Location	1828.953	609.651	1.189	0.316
Pavement type	414.235	414.235	0.804	0.371
Scope	978.158	326.053	0.629	0.597
Road classification	1773.496	886.748	1.738	0.179
Median strip	143.232	143.232	0.277	0.599

From Table 1, majority of the road projects executed within the Ghanaian construction industry in the last five years are within the southern part of the country with approximately 31.9% of the total 166 road projects retrieved. Roads executed in the western part of Ghana totalled 26.5%, eastern roads are 25.3% while northern roads are 16.3% in total. Roads constructed without a median strip account for 69.9% while roads with a median strip account for 30.1% of the total. With only 9% of the roads being unpaved, 91% are paved with either asphalt surface or paving stones. Most of the road constructions carried out are renovation works that totalled 50.6%, while 34.9% are road projects upgraded by increasing the length and width with other features. Roads replaced with new surfacing material account for 7.8% and newly constructed roads account for 6.6%. Out of the 166 road projects retrieved for this analysis, 68.7% are feeder roads while the trunk and urban roads are 16.9% and 14.5%, respectively. The information above shows that feeder roads are the most constructed classification of roads in Ghana within the last five years. Due to these roads being feeder roads, most of them do not have a median strip as they connect the trunk roads to the urban roads. A larger number of the roads executed are paved with asphalt surfacing material and are renovation works although some are upgraded to meet up prevailing standards as well. The cost performance of the road projects was calculated by deducting the budgeted construction cost from the final construction cost and finding the percentage of deviation from the budgeted construction cost as shown in eq. 3.

$$\text{Cost performance}\ (\%) = \left\{\frac{\text{Final cost} - \text{Budgeted cost}}{\text{Budgeted cost}}\right\} \times 100 \tag{3}$$

The cost performance calculation shows there are 122 cost overruns while 36 road projects experienced cost underruns and

Fig. 2. Scatterplot of cost performance against budgeted cost. [Colour online.]



8 road projects were completed within budget. The empirical data was therefore subjected to further analysis as shown in Table 2. The mean value for cost performance = 11.13%, cost overrun = 17.07%, and cost underrun = −6.50%. The standard deviation for cost performance = 22.69%, cost overrun = 23.67%, and cost underrun = 4.23%. The median values are 5.43%, 8.83%, and −7.24% for cost performance, cost overrun, and cost underrun, respectively.

The standard deviation indicates that most of the deviation recorded in cost performance and cost overruns are not close to the mean value, while the cost underruns recorded are close to the mean value. The kurtosis and skewness recorded showed that the data distribution is not at the extreme of the mean value and as such it is a normal distribution.

The Pearson's correlation analysis as shown in Table 3 indicated that only road width is statistically significant with cost performance at $p < 0.05$ (correlation value of 0.157) while project scope (new, renovation, upgrade, and replacement) is statistically significant with cost overruns at $p < 0.05$ (correlation value of 0.178) and road classification (trunk, feeder, and urban) is statistically significant with cost underruns at $p < 0.01$ (correlation value of 0.469). This shows that other road characteristics influence the deviation of the final construction cost of road projects from the budgeted construction cost in Ghana. This is contrary to the findings of Love et al. (2018) who observed that only the partners involved in the projects influence the deviation of final construction cost from the budgeted cost in the Australian construction industry.

The "one-way between-groups ANOVA" conducted and shown in Table 4 to examine the differences between the road project characteristics which are location of road projects, features of road projects (median strip, road length, road width, type of pavement), the scope of the road projects (new project, renovation work, upgrade work or replacement work), and classification of road projects (trunk, feeder, urban). However, road length and width were excluded from the analysis as they do not have uniform variables, unlike other project characteristics. "Levene's homogeneity of variance assumption" was not breached by the road project data distribution as all the characteristics have significance values above 0.05. The one-way ANOVA also showed that statistically significant differences do not exist at $p < 0.05$ level. This meant that there is no variance in the data distribution groups. There is, therefore, no need to carry out "Tukey's honest significant difference (HSD) post hoc test".

A scatterplot distribution of budgeted costs of road projects against the cost performance of the projects is shown in Fig. 2. The distribution indicated that there are more cost overruns experienced compared to cost underruns on executed road projects in Ghana.

Table 5. Goodness-of-fit test summary.

Distribution	Significance level (α)	Anderson Darling statistics	Kolmogorov–Smirnov statistics	Chi-Squared statistics
Cauchy cost performance (N = 166)	0.2	1.3749	0.08536	9.8032
	0.1	1.9286	0.0973	12.017
	0.05	2.5018	0.10804	14.067
	0.02	3.2892	0.12077	16.622
	0.01	3.9074	0.1296	18.475
Johnson S_B cost overruns (N = 122)	0.2	1.3749	0.0914	N/A
	0.1	1.9286	0.11073	N/A
	0.05	2.5018	0.12295	N/A
	0.02	3.2892	0.13743	N/A
	0.01	3.9074	0.14748	N/A
Johnson S_B cost underruns (N = 36)	0.2	1.3749	0.17418	5.9886
	0.1	1.9286	0.1991	7.7794
	0.05	2.5018	0.22119	9.4877
	0.02	3.2892	0.24732	11.668
	0.01	3.9074	0.26532	13.277

Information in Table 5 shows the goodness of fit summary generated from the analysis of the cost performance, overrun and underrun distribution. The table shows the best-fit probability distribution for the three categories using Kolmogorov–Smirnov statistics, Anderson Darling statistics, and Chi-Squared statistics at different significance levels. Cauchy distribution was found most suitable for cost performance while Johnson S_B was found suitable for both cost overruns and underruns.

Figures 3 and 4 indicate the distribution of the mean cost overrun and mean cost underrun as experienced on road projects based on the project sizes. Projects <\$900 000.00 experienced 15.15% and −3.61% mean cost overrun and underrun respectively; projects >\$900 000.00 had 9.23% and −7.08% mean cost overrun and underrun respectively; projects >\$2700 000.00 had 33.04% and −0.72% mean cost overrun and underrun respectively; projects >\$7200 000.00 had 12.72% and −8.28% mean cost overrun and underrun respectively; projects >\$12 600 000.00 had mean cost overrun and underrun of 51.58% and −6.22% respectively; while projects >\$21 600 000.00 had 6.02% and −0.17% mean cost overrun and underrun respectively.

The discrete probabilities of cost overruns and underruns on the budgeted construction cost for road projects in Ghana are shown in Table 6. Johnson S_B indicates the categorization of the

Fig. 3. Distribution of cost overrun according to project size. [Colour online.]

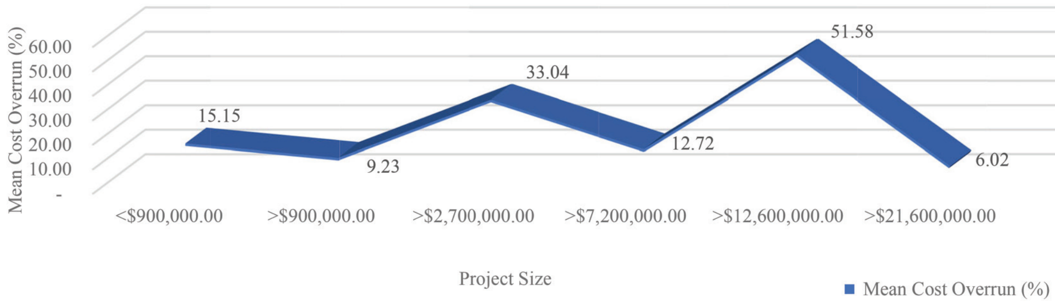


Fig. 4. Distribution of cost underrun according to project size. [Colour online.]

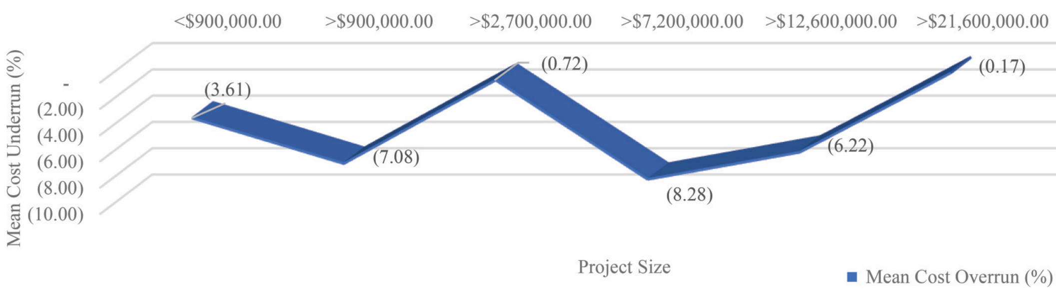


Table 6. Discrete probabilities of cost overruns and underruns.

Distribution	Probabilities	P(X1 < X < X2)
Johnson S_B cost overruns (N = 122)	1 and 20%	0.82
	21 and 40%	0.04
	41 and 60%	0.04
	61 and 80%	0.03
	81 and 100%	0.02
Johnson S_B cost underruns (N = 36)	–1 and –3%	0.25
	–4 and –6%	0.14
	–7 and –9%	0.21
	–10 and –12%	0.14
	–13 and –15%	0.05

probabilities based on percentages. Cost overruns were examined between 1% and 100% while cost underruns were examined between −1% and −15%.

Discussion

Cost performance

Using the Anderson Darling statistics, Kolmogorov–Smirnov statistics, and Chi-Squared statistics goodness-of-fit to determine the best-fit among the probability distributions for cost performance of road projects, Cauchy distribution was found as the most suitable with parameters σ = 6.21 and μ = 5.57. Cauchy distribution mostly referred to Lorentz distribution is a continuous probability distribution. It has been adjudged to be one of the few probability distributions with stability attribute and expressible analytically according to NIST (NIST Sematech 2019). The Anderson Darling statistics = 2.2758, Kolmogorov–Smirnov statistics = 0.09293, and Chi-Squared statistics = 10.119. The significance at different p values show there is no statistical significance for each of the critical values under each goodness-of-fit tests. It can, therefore, be said that σ is an uninterrupted scale parameter (where $\sigma > 0$) while μ is denoted as an uninterrupted location parameter. From the foregoing, $-\infty < x < +\infty$ is the domain for this distribution. Figures 5 and 6 indicated the cost performance using Cauchy cumulative density function and probability density function. The histogram distribution shows the different levels of cost deviation in percentages while the curve shows the distribution of the deviations. The graph indicates the percentage of cost performance against the probability of cost performance on road projects. The cost performance probabilities using the Cauchy CDF and PDF functions are defined in the following equations respectively.

$$F(x) = \left\{ \pi\sigma \left[1 + \left(\frac{x-\mu}{\sigma} \right)^2 \right] \right\} \tag{4}$$

$$F(x) = \frac{1}{\pi} \arctan\left(\frac{x-\mu}{\sigma} \right) + 0.5 \tag{5}$$

Cost overruns

Using the Anderson Darling statistics, Kolmogorov–Smirnov statistics, and Chi-Squared statistics goodness-of-fit also to determine the best-fit among the probability distributions for cost performance of road projects, Johnson S_B distribution was found as the most suitable with parameters γ = 1.5779, δ = 0.43844, λ = 131.38, ξ = 3.1242. Johnson S_B distribution according to Love et al. (2018) is a four-parameter lognormal model that has two properties which makes it suitable for a sample of projects' cost overruns and underruns distribution. The first property shows upper and lower bound $\xi + \lambda$ and ξ respectively which can represent financial constraints by financiers of the road projects. The second property indicates shape parameters δ and γ gives room for flexibility to allow a large distribution of data. The Anderson Darling statistics = 77.689 and Kolmogorov–Smirnov statistics = 0.17038. Chi-Squared statistics was not generated for the cost overrun distribution to indicate that the data distribution is not evenly distributed. It can, therefore, be said that ξ is the continuous location parameter, λ is the continuous scale parameter (where $\lambda > 0$) while δ and γ ($\delta > 0$) are denoted as the continuous shape parameters. From the above, $\xi \le x \le \xi + \lambda$ is the domain for this distribution. The probability of experiencing cost overrun between

Fig. 5. Cauchy CDF for cost performance. [Colour online.]

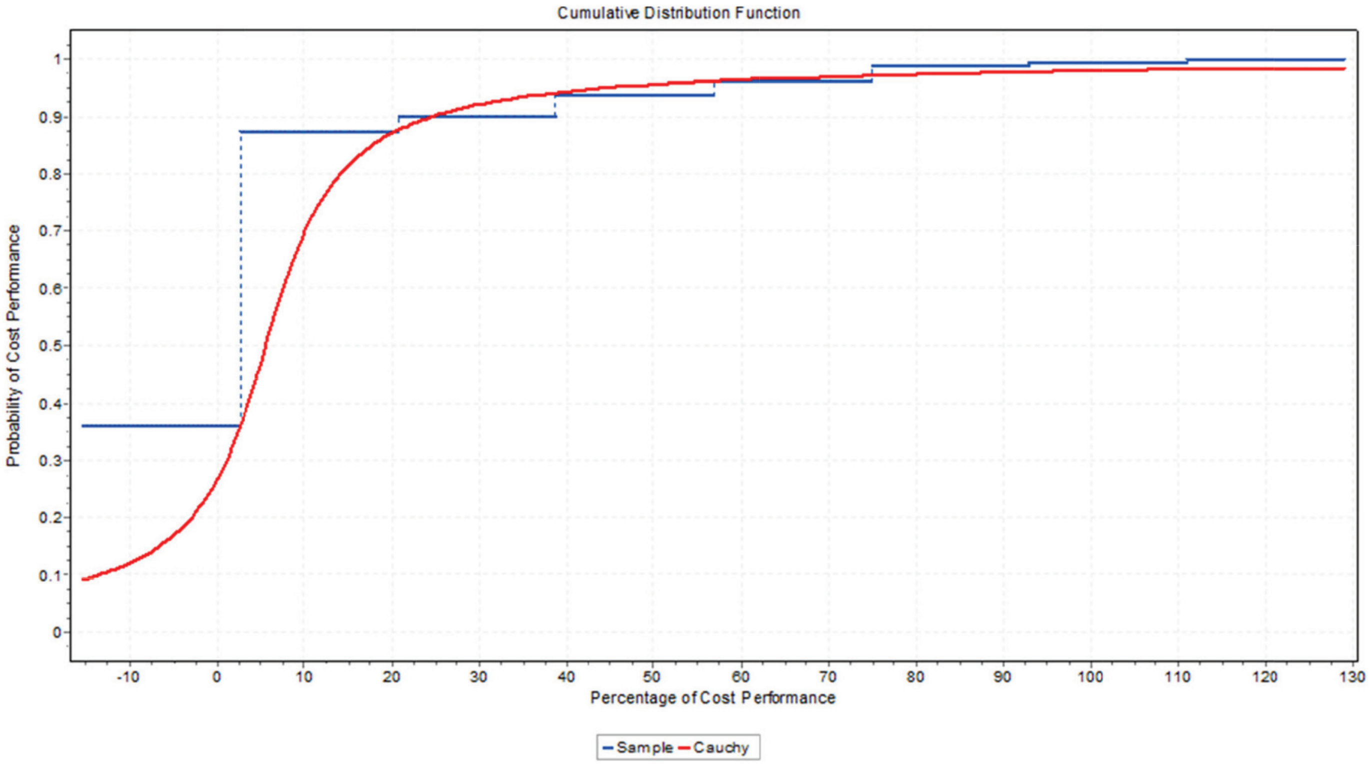


Fig. 6. Cauchy PDF for cost performance. [Colour online.]

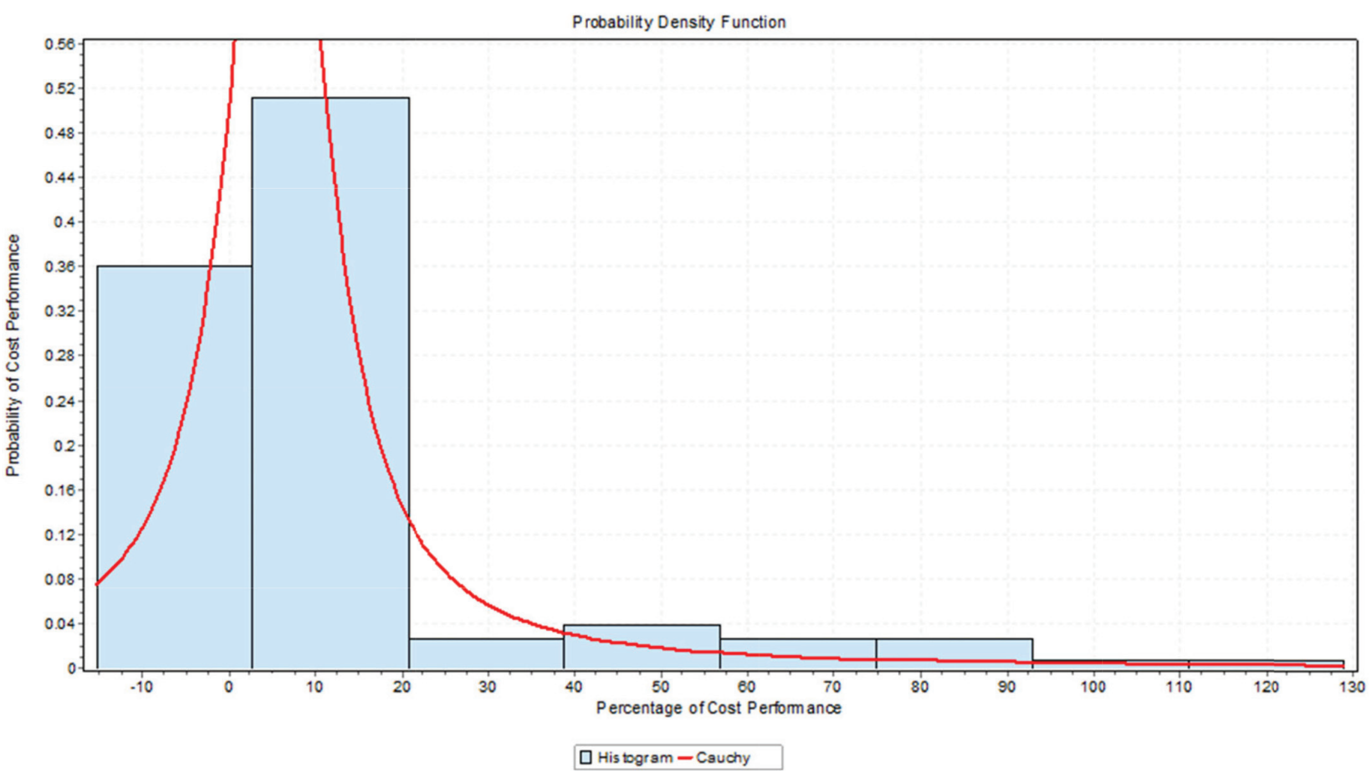
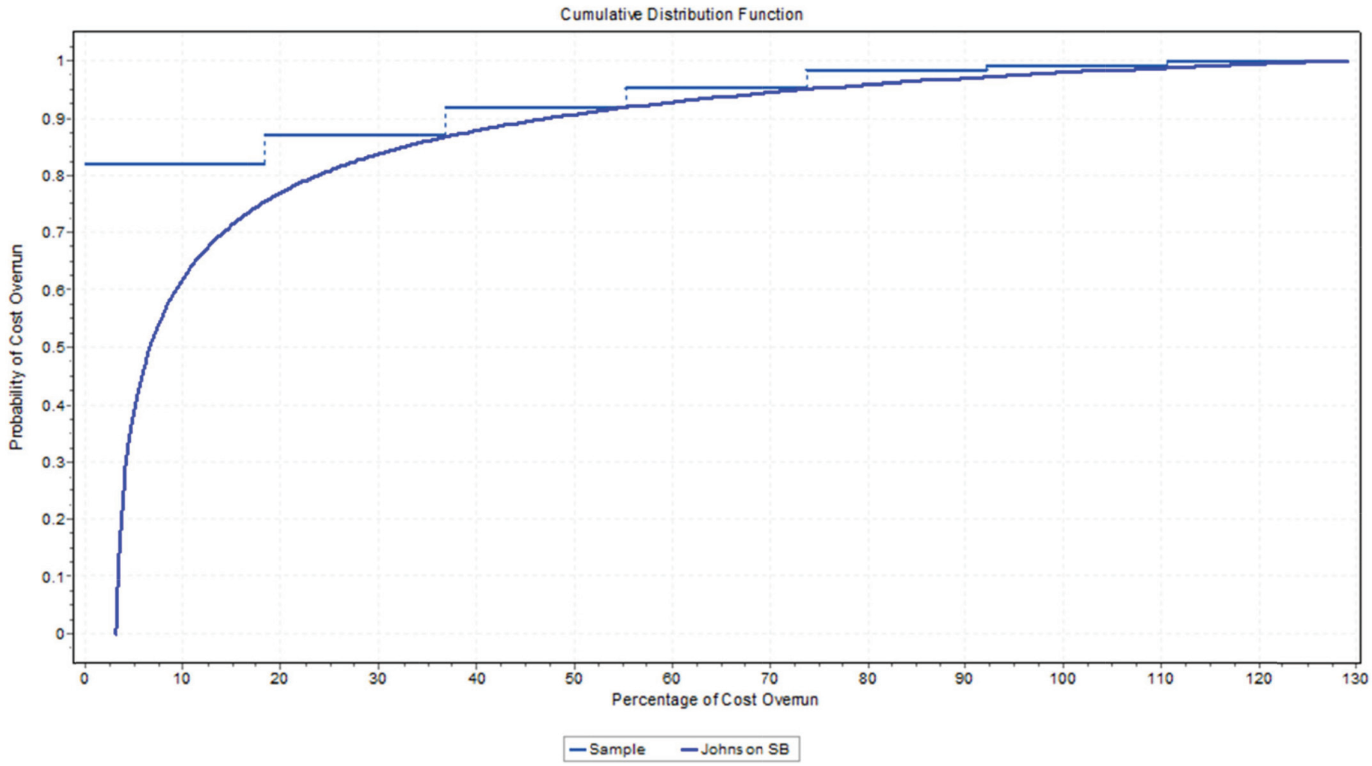


Fig. 7. Johnson S_B CDF for cost overrun. [Colour online.]



1% and 20% over the budgeted construction cost is 82%; cost overrun between 21% and 40%, 41% and 60% is both 4%; cost overrun between 61% and 80% is 3% while cost overrun between 81% and 100% is 2%. This is contrary to the study of Love et al. (2015) who found that Log-logistic (3 P) distribution is more suitable for the cost overrun experienced on road projects in Australia. However, Love et al. (2018) agreed that Johnson S_B distribution is suitable for cost overrun on water infrastructure. The cost overrun probabilities using the Johnson S_B PDF and CDF functions is defined in the following equations respectively.

$$F(x) = \frac{\delta}{\lambda\sqrt{2\pi z(1-z)}}\exp\left\{-\frac{1}{2}\left[\gamma + \delta \text{In}\left(\frac{z}{1-z}\right)\right]^2\right\} \tag{6}$$

$$F(x) = \Phi\left[\gamma + \delta \text{In}\left(\frac{z}{1-z}\right)\right] \tag{7}$$

where $z = (x - \xi)/\lambda$ and Φ = Laplace integral.

The cost overrun of the historical data distribution was generated using Johnson S_B cumulative density function and probability density function as indicated in Figs. 7 and 8, respectively. The histogram distribution shows the different levels of cost overrun experienced on road projects in percentages, while the curve shows the distribution of the overruns. The graph indicates the cost overrun percentage against the cost overrun probability on road projects.

Cost underruns

Using the Anderson Darling statistics, Kolmogorov–Smirnov statistics, and Chi-Squared statistics goodness-of-fit also to determine the best-fit among the probability distributions for cost performance of road projects, Johnson S_B distribution was found as the most suitable with parameters γ = 0.11394, δ = 1.1577, λ = 22.684, ξ = −18.318. This contradicts the study of Love et al. (2015) who observed that Wakeby distribution is more suitable for cost underrun experienced on water infrastructure in the UK. The Anderson Darling statistics = 0.57927, Kolmogorov–Smirnov statistics = 0.14999, and Chi-Squared statistics = 11.198. Thus, goodness-of-fit tests reveal that the data distribution is evenly distributed. The significance at different p values show there is no statistical significance for each of the critical values. As stated earlier for Johnson S_B distribution, ξ is the continuous location parameter, λ is the continuous scale parameter (where $\lambda > 0$) while δ and γ ($\delta > 0$) are denoted as the continuous shape parameters. The domain for this distribution is $\xi \leq x \leq \xi + \lambda$. The probability of having a cost underrun of the budgeted construction cost between −1% and −3% according to the best-fit distribution is 25%; cost underrun between −4% and −6% is 14%; cost underrun between −7% and −9% is 21%; cost underrun between −10% and −12% is 14%; cost underrun between −13% and −15% is 5%. The Johnson S_B CDF and PDF are defined in eqs. 6 and 7. Cost underrun of the historical data distribution was generated using Johnson S_B cumulative density function and probability density function also as indicated in Figs. 9 and 10, respectively. The histogram distribution shows the different levels of cost underrun experienced on road projects in percentages, while the curve shows the distribution of the underruns. The graph indicates the cost overrun percentage against the cost overrun probability on road projects.

Based on the findings from the analysis carried out, cost overrun occurs more on road construction projects in Ghana compared to cost underrun, while it is evident that only a few projects are achieved within the budgeted construction cost. Projects within \$12 600 000.00 and \$21 600 000.00 budgeted construction cost experience cost overrun more than other categories of road projects, while projects within \$7 200 000.00 and \$12 600 000.00 experience more cost underrun than other categories. With the separation of cost overrun and underrun, the likelihood of occurrence was independently analyzed to give improved risk management on road projects. Road projects are more likely to experience cost overrun between 1% and 20% as it was indicated

Fig. 8. Johnson S_B PDF for cost overrun. [Colour online.]

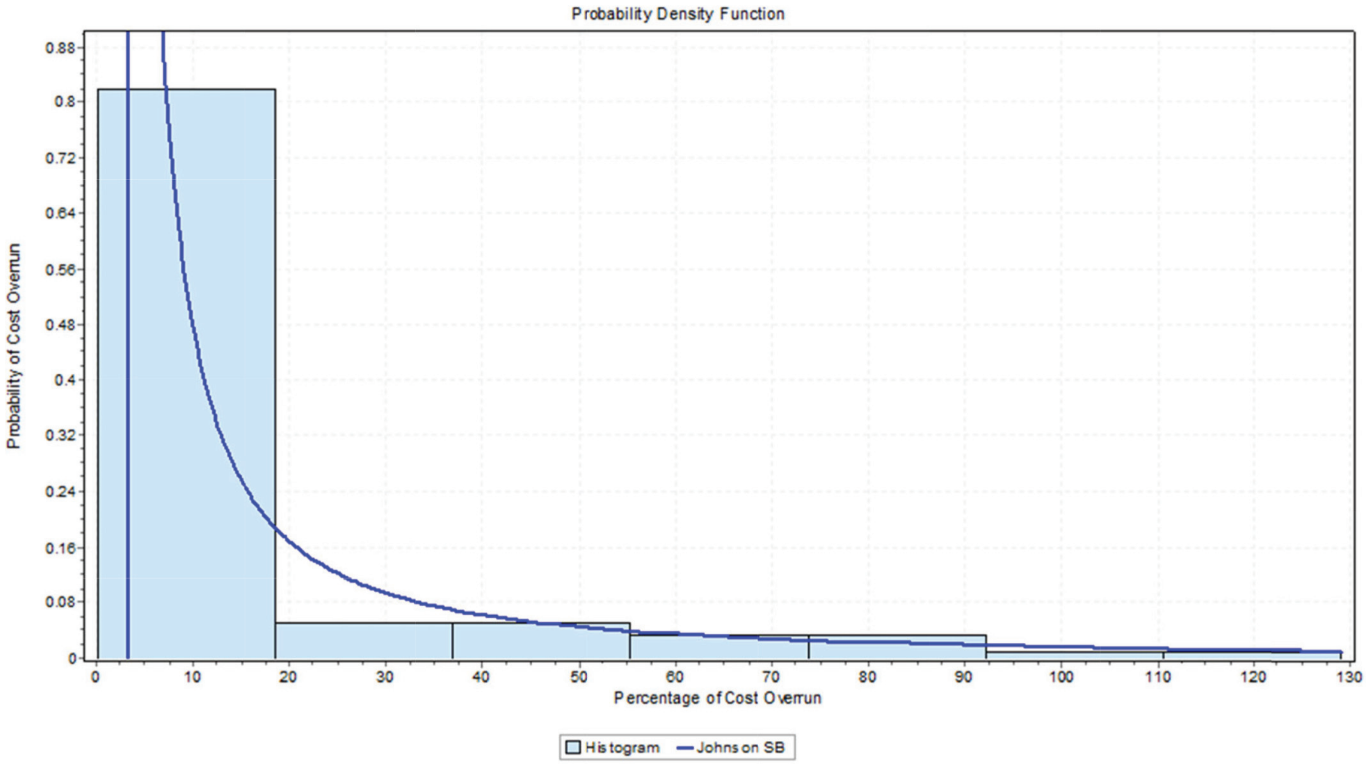


Fig. 9. Johnson S_B CDF for cost underrun. [Colour online.]

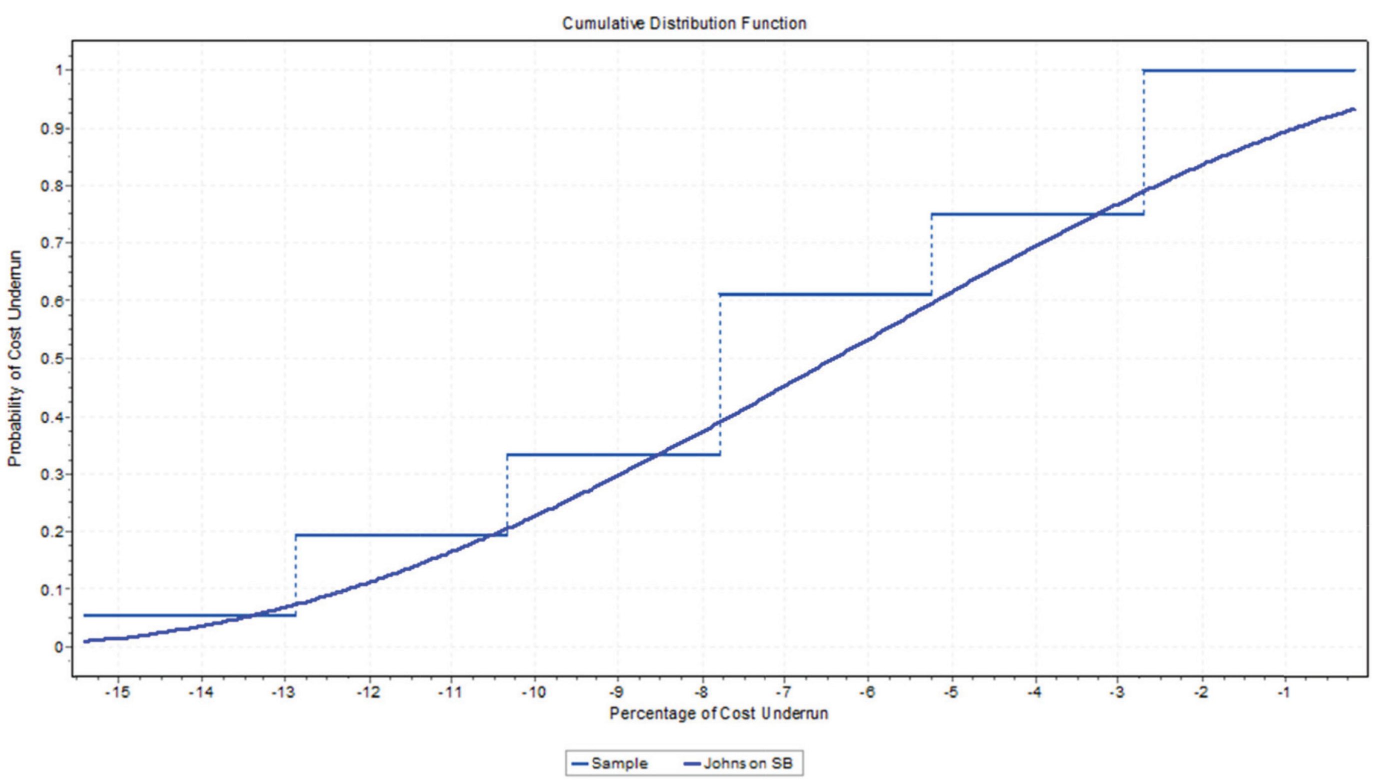


Fig. 10. Johnson S_B PDF for cost underrun. [Colour online.]

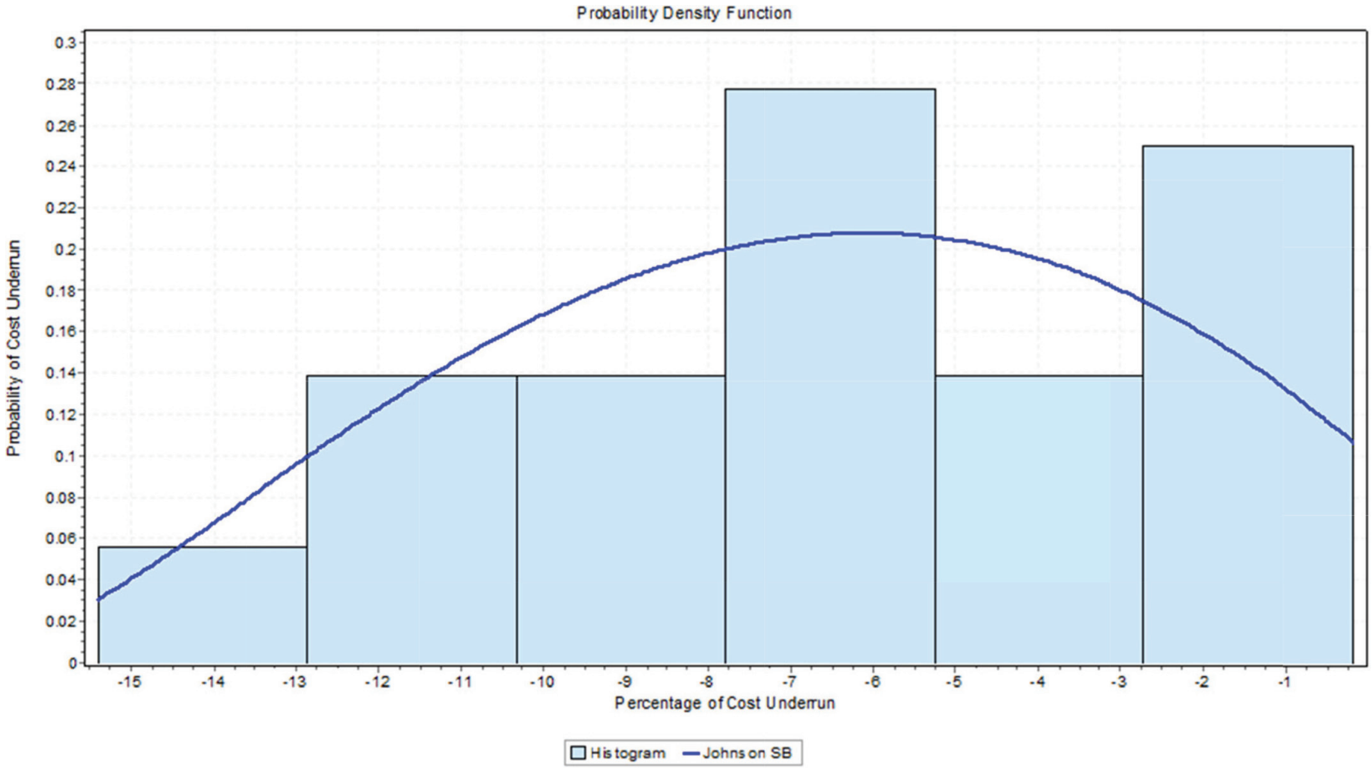


Table 7. Validation of probability distribution.

S/N	Budgeted construction cost ($)	Final construction cost ($)	Cost deviation ($)	Cost performance (%)
1	20 955 129.74	20 955 129.74	—	—
2	16 501 384.12	18 164 348.20	1 662 964.08	10.08
3	43 392 812.80	45 877 950.25	2 485 137.45	5.73
4	2 241 048.23	2 555 421.63	314 373.40	14.03
5	2 387 374.52	2 415 623.90	28 249.38	1.18
6	41 211 367.91	48 098 387.73	6 887 019.82	16.71
7	41 600 256.50	44 542 970.04	2 942 713.54	7.07
8	156 999 494.83	184 945 478.30	27 945 983.47	17.80
9	73 545 247.75	84 554 650.97	11 009 403.22	14.97
10	148 578 336.95	146 450 755.78	−2 127 581.17	−1.43

that there is an 82% probability of such occurrence. Other than this, the probability of experiencing cost underrun between −1% and −12% is approximately 19%. To validate the probability distribution, executed projects which are not used in the cost profiling was examined in Table 7.

From the result presented in the table, eight projects experienced cost overrun in which all the overruns are between 1% and 20% which confirms the probability distribution that stated that there is a 82% likelihood of 1% to 20% cost overrun on road projects executed in Ghana. The only cost underrun experienced is between −1% and −3% in accordance with the probability distribution generated from the cost profiling carried out. This, therefore, shows that the objective was achieved.

Limitations

The study encountered some distinct limitations notably among all is that the road projects considered for the cost profiling are projects financed by the government in Ghana. Government-financed projects are known to experience various delays which are major factors contributing to cost overrun on construction projects. Due to this reason, private-financed projects cannot rely on the probability output of this study to determine the accuracy of its estimates.

Conclusion and recommendations

Analyzed data revealed that the majority of road projects executed in Ghana experience cost overrun while some also experience cost underrun. However, only a few projects are achieved within the budgeted construction cost with a mean cost deviation of 22.69%. Among the identified seven project characteristics namely project location, length, width, inclusion of median strip, pavement type, scope, and road classification, only three characteristics influence the cost profile of road projects in Ghana which are road width, scope, and road classification. These characteristics play a major role in the output of the cost probabilities. Estimates prepared for road projects can be expected to be below the final construction cost by approximately 20%. This is evident from records of executed road projects as most of the roads experience cost overrun more than cost underrun. On average, estimates for road projects in Ghana are inaccurate as only very few projects are completed within the budgeted construction cost. With the various challenges to the accuracy of estimates for construction projects affecting the cost performance of the construction industry, it is recommended that this cost profile for government-funded road projects in Ghana be employed. The cost profile revealed that an allowance of 20% for contingency on road project estimate is adequate to avoid cost overrun on road projects in Ghana. Also, to ensure estimates for road projects in Ghana are accurate as expected, risks must be adequately evaluated while contingency needs to be incorporated into the estimated cost of road projects. However, proper checks must be put in place to ensure contingency sums are expended rightly to avoid fraudulent and sharp practices by consultants and contractors.

Acknowledgement

The authors would like to acknowledge the financial support provided by the University of Johannesburg (Global Excellence Stature) and also acknowledge the two peer reviewers who committed their time into reviewing this article. On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Canhoto, A., and Backhouse, J. 2008. General description of the process of behavioural profiling. *In* Profiling the European citizen. Springer, Dordrecht, the Netherlands, pp. 47–63. doi:10.1007/978-1-4020-6914-7_3.

Chileshe, N., and Berko, P.D. 2010. Causes of project cost overruns within the Ghanaian road construction sector. *In* Proceedings of the 5th Built Environment Conference (ASOCSA), Durban, South Africa, pp. 66–81.

Chou, J.S., Wang, L., Chong, W.K., and O'Connor, J.T. 2005. Preliminary cost estimates using probabilistic simulation for highway bridge replacement projects. *In* Construction Research Congress 2005: Broadening Perspectives - Proceedings of the Congress, pp. 939–948. doi:10.1061/40754(183)92.

Edens, J.F. 2001. Misuses of the hare psychopathy checklist-revised in court. Journal of Interpersonal Violence, **16**(10): 1082–1093. doi:10.1177/088626001016010007.

Flyvbjerg, B. 2007. Cost overruns and demand shortfalls in urban rail and other infrastructure. Transportation Planning and Technology, **30**(1): 9–30. doi:10.1080/03081060701207938.

Hegazy, T. 2002. Computer-based construction project management. Prentice-Hall, Upper Saddle River, N.J., pp. 476–479. doi:10.1016/B978-0-12-408090-4.00008-6.

Hildebrandt, M., and Gutwirth, S. 2008. Defining profiling: A new type of knowledge? *In* Profiling the European citizen: cross-disciplinary perspectives. Springer, Dordrecht, the Netherlands, pp. 17–45. doi:10.1007/978-1-4020-6914-7_2.

Kim, G.H., An, S.H., and Kang, K.I. 2004. Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. Building and Environment, **39**(10): 1235–1242. doi:10.1016/j.buildenv.2004.02.013.

Liu, L., and Zhu, K. 2007. Improving cost estimates of construction projects using phased cost factors. Journal of Construction Engineering and Management, **133**(1): 91–95. doi:10.1061/(ASCE)0733-9364(2007)133:1(91).

Love, P.E.D., Wang, X., Sing, C.-P., and Tiong, R.L.K. 2013. Determining the probability of cost overruns in Australian construction and engineering projects. Journal of Construction Engineering and Management, **139**(3): 321–330. doi:10.1061/(ASCE)CO.1943-7862.0000575.

Love, P.E.D., Sing, C.-P., Carey, B., and Kim, J.T. 2015. Estimating construction contingency: accommodating the potential for cost overruns in road construction projects. Journal of Infrastructure Systems, **21**(2): 04014035. doi:10.1061/(ASCE)IS.1943-555X.0000221.

Love, P.E.D., Ahiaga-Dagbui, D.D., and Irani, Z. 2016. Cost overruns in transportation infrastructure projects: Sowing the seeds for a probabilistic theory of causation. Transportation Research Part A: Policy and Practice, **92**: 184–194. doi:10.1016/j.tra.2016.08.007.

Love, P.E.D., Ahiaga-Dagbui, D.D., Smith, S.D., Sing, M.C.-P., and Tokede, O. 2018. Cost profiling of water infrastructure projects. Journal of Infrastructure Systems, **24**(4): 04018023. doi:10.1061/(ASCE)IS.1943-555X.0000441.

NIST Sematech. 2019. Engineering statistics handbook, exploratory data analysis. Available from https://www.itl.nist.gov/div898/handbook/eda/section3/eda3663.htm [accessed 14 August 2019].

Ofori-Kuragu, J.K., Owusu-Manu, D.G., and Ayarkwa, J. 2016. The case for a construction industry council in Ghana. Journal of Construction in Developing Countries, **21**(2): 131–149. doi:10.21315/jcdc2016.21.2.7.

Ogungbile, A.J., Oke, A.E., and Rasak, K. 2018. Developing cost model for preliminary estimate of road projects in Nigeria. International Journal of Sustainable Real Estate and Construction Economics, **1**(2): 182–199. doi:10.1504/IJSRECE.2018.092277.

Osei, V. 2013. The construction industry and its linkages to the Ghanaian economy-polices to improve the sector's performance. International Journal of Development and Economic Sustainability, **1**(1): 56–72.

Rossi, M.E., and Deutsch, C.V. 2014. Statistical tools and concepts. *In* Mineral resource estimation. Springer, Dordrecht, the Netherlands, pp. 11–27. doi:10.1007/978-1-4020-5717-5_2.

Seneviratne, S., and Levy, D.C. 2010. Cost profile prediction for grid computing. Concurrency Computation Practice and Experience, **22**(1): 107–142. doi:10.1002/cpe.1537.

Swei, O., Gregory, J., and Kirchain, R. 2017. Construction cost estimation: A parametric approach for better estimates of expected cost and variation. Transportation Research Part B: Methodological, **101**: 295–305. doi:10.1016/j.trb.2017.04.013.