CITC GLOBAL
Construction in the 21st Century
11th International Conference (CITC-11)
September 9-11, 2019, London, UK

Springer
SPRINGER NATURE

Modelling correlations in highway construction projects

Alireza Moghayedi[1] and Abimbola Windapo[1]

[1]University of Cape Town, Cape Town, South Africa
mghali001@myuct.ac.za

Abstract. The study examines the different types of correlations between the construction costs, times and cost/time of highway construction projects and quantified their impacts on the total cost and total time of different structures of highway construction projects to determine whether repetitions of activities amplify the impact of correlation on construction cost and time. Highway construction projects are often plagued by cost and time underestimation due to ignorance of costs and times correlation between the activities in highway projects when deterministic estimation techniques are employed. Therefore, a probabilistic model is adopted in estimating the effect of correlations on the probability distributions of total cost and total time. This study identified various types of correlations between the costs, times and cost/time in the construction of highway projects and modelled them using Gaussian copula and analyzed the impact of such correlations in the construction of a highway project using Monte Carlo simulation. The results show that the standard deviation of the total cost and total time increases with the magnitude of the correlation and type of correlation matrix and, most importantly, it increases considerably with the number of costs and times that are correlated. Based on these findings, the study concludes that the deterministic estimation technique in use does not capture the wide range of the possible total cost and time of highway projects resulting in significant overruns.

Keywords: correlation, cost and time underestimation, cost and time overruns, deterministic estimation, highway construction, repeated activity

1 Introduction

Transportation infrastructure projects, particularly highway construction projects, are often plagued by cost and time underestimations [1]. Various studies [2-5] proved that both construction costs and times are positively correlated. Positive correlations caused the standard deviation of the correlated costs and times to be larger than the standard deviation of uncorrelated costs and times, and the total construction cost and time of projects were underestimated if the costs, times and cost/time correlation were disregarded [2].

Correlation measures and describes the strength and direction of the relationship between two variables [6]. The correlation varies between -1 (fully negatively correlated) and +1 (fully positively correlated), and, for a correlation equal to 0, the two variables are uncorrelated. If the value of one variable is above average, the value of the second variable tends to be above average when they are positively correlated, while it tends to be below average when they are negatively correlated [7].

A positive correlation causes the standard deviation of the sum of the correlated variables to increase since cost and time correlations in construction are usually positive [2, 4]. Compared to deterministic estimated total cost and total time, the correlated cost and time are expected to increase the range of possible total costs and total times of the project, which generates the cloud of points in the total cost and total time directions.

Differently from the deterministic cost and time estimate, which corresponds to one point, modelling the correlation of costs and times provides the range of the possible total cost and total time by a cloud of points. The cloud of points visually represents the range of possible correlated total costs and total times of the different structures of the project.

Correlation between the costs and times has been investigated in building construction [3] and rail line construction [4], but there is a lack of studies on correlation in highway construction projects. Therefore, this study examines various types of correlation that occur in a highway construction project and analysed their impact on the distributions of the cost and time of three main structures of the highway construction project, namely: earthwork, bridge, and pavement, to determine whether repetitions of activities amplify the impact of correlations on such distributions.

2 Methodology

Correlations in the construction of a highway are analyzed based on the following methodology steps. First, various types of correlation in highway construction are identified. After that, the correlation coefficient of each type of correlation is determined from the analysis of historical data by a panel of experts (which, in the study, comprised five highway experts with experience of 30 years and more individually in the construction of highway projects in South Africa recommended by the South African national roads agency). Identified types of correlation are modeled and simulated in three structures of highway construction. Finally, the simulation results are analysed by comparing mean and standard deviation to the independent results; and findings and conclusions are discussed.

3 Correlations in the construction of highway projects

Correlations in the construction of highway projects are investigated on the basis of a construction network model of a highway. In the network model, the construction of a highway project is modelled as the construction of a sequence of three main structures, namely, earthwork, bridge and pavement, each of which is modeled as a sequence of various construction activities characterized by cost and time.

The study identified the following five potential types of correlation in the construction of highway projects:

Correlation between the costs and times of a repeated activity in a structure (Type 1): The costs and times of a repeated activity are expected to be positively correlated because of the repetitiveness of the processes in a structure [2]. To analyse the impact of repeated activities on the cost and time of activities of each highway structure, the intermediate correlation matrix including all positive correlations, ρ = (0, 1), is considered, which corresponds to the case where the cost and time are randomly selected for each unit, and if one cost or time per unit is above average, the next cost or time per unit will also tend to be above average.

$$\rho = \begin{bmatrix} 1 & 0.99 & \dots & 0 \\ 0.99 & 1 & 0.99 & \vdots \\ \vdots & 0.99 & 1 & 0.99 \\ 0 & \dots & 0.99 & 1 \end{bmatrix}$$

Correlation between the costs and times of the various activities in a structure (Type 2): To evaluate the correlation between the various activities in each of the three structures of the highway project, the expert panel of the study considered that the costs and times of the various activities in a structure were positively correlated because these activities were subjected to the same types of constraints. After analysis of the historical data of highway construction projects, the panel of experts recommended the correlation coefficients of ρ =0.75, ρ =0.65 and ρ =0.8 for earthwork, bridge and pavement structures, respectively.

Correlation between the costs and times of activities in adjacent structures (Type 3): Although the pairs structures of earthwork-pavement and bridge-pavement are adjacent structures, they do not share any activities. Therefore, the expert panel recommended no correlation between the costs and times of these pairs. Earthwork and the bridge are excavated in the same geology so that a correlation between the costs and times of the cutting in earthwork and the cost and time of the bridge excavation is expected. However, due to the different excavation methods between the cutting and bridge excavation, the panel of experts recommended that the cost and time of earthwork and bridge should be independent.

Correlation between the costs and times of same activities in the same types of structures (Type 4): Positive correlations between the costs and times of the same types of structures (between bridge i and bridge j, between earthwork i and earthwork j, and between pavement i and pavement j) were expected if the geology (earthwork, bridge) and geometry (bridge) were similar. The cost and time distribution of bridge and earthwork structures were a function of the geology and the cost and time of a bridge were a function of the geometry so that similar geologies and geometry determined similar cost and time distributions.

Correlation between the cost and time of an activity (Type 5): Due to the difficulty of measuring the association between the cost and time of an activity, all members of the expert panel agreed to use ρ=+0.8 as correlation coefficient to model the correlation between the cost and time of activities.

$$\rho = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$$

Therefore, the study focused on modelling correlation Type 1 (correlation between repeated activities), correlation Type 2 (correlation between different activities), and correlation Type 5 (correlation between cost and time of the activity), which was expected to have a significant effect on the standard deviation and the total cost and time of construction of highway projects.

4 Modelling correlations

In this study, the correlation was measured with the non-parametric Spearman correlation coefficient and modelled with the Gaussian copula. A copula is a multivariate distribution function defined on the unit cube $[0, 1]^d$, with uniformly distributed marginals [8].

The choice of the Spearman correlation from amongst the available correlation measures and the Gaussian copula from amongst the available correlation models was because, differently from a random number generator, a copula first generates random numbers from a uniform distribution, then, through two transformations, it obtains random numbers correlated with the desired correlations and distributed with the desired probability distribution. Also, a copula generates the unit costs and production rates of the construction activities, which are correlated, then these are summed to obtain the total cost and total time of the activities. In other words, copulas allow the generation of correlated costs and times with the desired Spearman correlation matrix and with the desired marginal distribution, and, because of the availability of both marginal distributions and the correlation matrix for this study, the copula was employed.

The Gaussian copula generation and the summing of the unit costs and production rates of all activities were repeated for each simulation run. With the copula generation, the uncertainty in each unit of construction activities was modelled, as well as the correlation between the costs and times of the activities. At the end of each simulation run, the costs and the times (critical activities), were respectively summed up to calculate the correlated total cost and total time of activities for the particular simulation run. The correlation process was repeated for every simulation run by Monte Carlo until the standard deviation of total cost and total time of activity bound within ±1% in 10 out of 10 simulations.

A correlated probabilistic analysis was required to model the correlations on the construction cost and time of activities and evaluate the impact of the identified three correlations on the total cost and total time of the project. Essentially, the probabilistic distributions data of the cost and time variables were not available. Therefore, to model the correlation with copula, the following marginal distributions assumptions were constructed by the panel of experts based on the results of the variability of cost and time estimation sessions to generate a correlation probability distribution of costs and times:

- The underlying distributions of cost variables and time variables are lognormal and triangular, respectively [3].
- The modes of the cost and time distribution are assumed equal to the deterministic cost and time.
- The minimum value of the cost and time distributions are 80% of the mode of each variable's distribution.
- There is a probability of 2% of exceeding the High Value (which is assumed to be 150% of the mode) of the cost distributions.
- The maximum of the time distributions is assumed to be 130% of the mode of the time distributions.

5 Results of the study

The deterministic estimated cost and time, and resulting means, standard deviations and variation of standard deviations of each type of correlation to the uncorrelated case (base case) from modelling the three types of correlation on total cost and total time of the earthwork, bridge and pavement structures of the selected highway construction project are summarized in Table 1.

The selected case study was a 16.5-kilometre section from a new highway infrastructure project with six lanes and controlled-access, which was constructed on both flat and hilly terrain in South Africa between 2014 and 2016. The deterministic estimated total construction cost of this section was ZAR 809,504,922.80 with a total duration of 340 days. The construction of the project comprised 2,973,692 cubic metres of earthworks in four different geological conditions, namely, fine, soft, hard and rock, 46 precast concrete bridges and 1,015,340 cubic metres of asphalt pavements.

Table 1. Correlated means, standard deviations and standard deviation variations of three structures of the highway project

	Correlation	Cost (ZAR)				Time (Days)			
		Deterministic	Mean	Standard Deviation	Δ Sd	Deterministic	Mean	Standard Deviation	Δ Sd
Earthwork Structure	Base Case	244,292,292.83	268726606.90	1871424.00	0.00%	215	222.19	2.61	0.00%
	Type 1		268180187.90	22538485.66	1104.35%		222.678	22.34	756.05%
	Type 2		268608078.40	14106038.00	653.76%		222.15	22.22	751.45%
	Type 3		268784679.2	18319948.86	878.93%		222.32	19.53	648.51%
Bridge Structure	Base Case	70,952,700.00	78,030,393.34	1,920,505.95	0.00%	340	351.35	5.04	0.00%
	Type 1		78,113,067.80	6,014,085.33	213.15%		351.36	30.15	498.24%
	Type 2		77,890,401.00	6,425,163.00	234.56%		351.46	32.81	551.42%
	Type 3		78,072,569.32	6,191,307.31	222.38%		351.57	22.61	348.55%
Pavement Structure	Base Case	494,259,930.00	543,664,957.99	4,174,001.38	0.00%	217	224.22	2.64	0.00%
	Type 1		543,853,187.64	52,912,363.32	1167.67%		224.27	13.23	401.14%
	Type 2		543,899,815.54	33,572,597.99	704.33%		224.41	11.13	321.59%
	Type 3		542,662,510.48	41,474,779.05	893.65%		223.93	11.59	339.02%

Furthermore, the cost and time of base case and correlation Type 1, 2 and 3 of earthwork, bridge and pavement structures from Monte Carlo simulations overlaid in scatterplots and are illustrated in Figures 1 to 3.

Cost correlations represent both threat and opportunity aspects. In fact, the correlations cause the range of the total cost and time on both sides of the mean, which is the threat of a total cost or time more substantial than the mean as well as the prospect of a total cost or time smaller than the mean total cost or time as illustrated in three overlaid scatterplots (Figs 1-3).

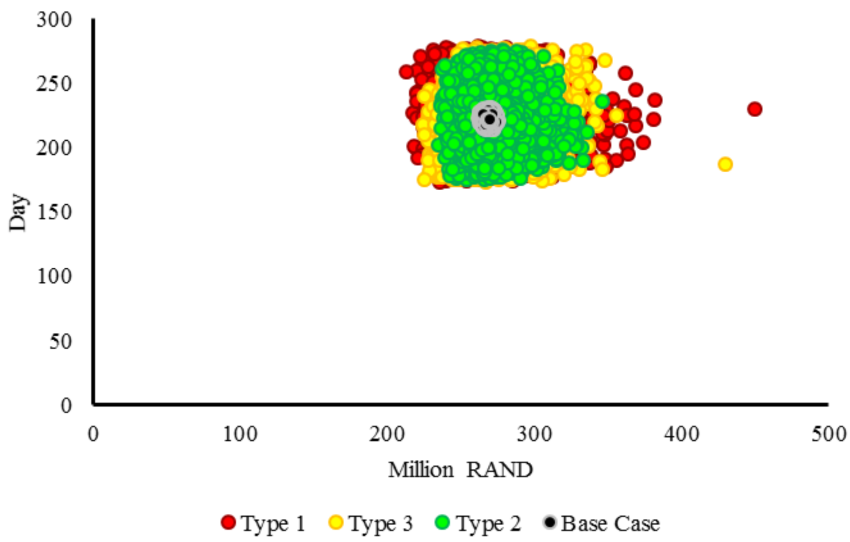

Fig. 1. Overlay scatterplot correlated cost and time of earthwork structure

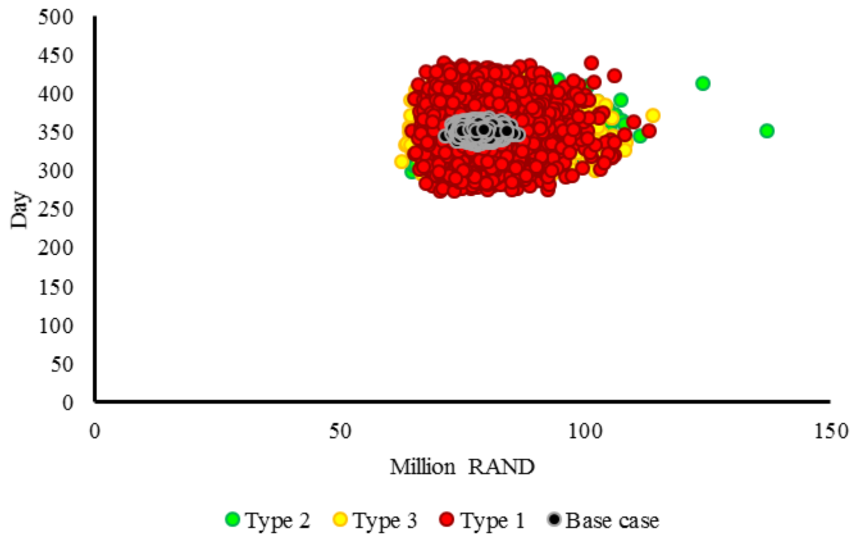

Fig. 2. Overlay scatterplot correlated cost and time of bridge structure

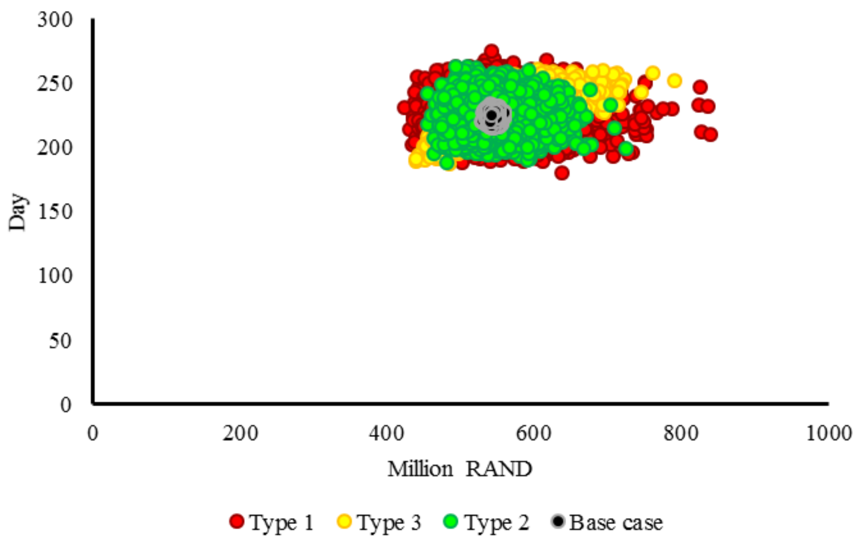

Fig. 3. Overlay scatterplot correlated cost and time of pavement structure

Based on the overlaying scatterplots of three types of correlation in the different structures of the highway project (Figs 1-3), and the results of means, standard deviations and variation of standard deviations of correlated cost and total time of the three structures of the highway project the following observations were made:

- The mean of the correlated cost and time are constant across different types of correlations (Table 1), which is confirmed by a constant centre of gravity in the cost and time-frequency plots (Figs. 1-3). This was expected because correlations affect the standard deviation and not the mean.
- The correlated means of total cost and total time are more significant than the deterministic estimated total cost and total time (Table 1) because the deterministic total cost and total time are the sums of the modes of costs and critical times of activities, but the correlated means of cost and total time are equal to the sum of the means of costs and critical times of activities; in other words, correlated cost and time are skewed to the right, (the mode is smaller than the mean) (Figs 1-3).
- The standard deviations of the cost and total time were increased considerably from base case (uncorrelated) to other correlation types, as shown in Table 3. The increases in total cost standard deviation and total time standard deviation are visible in the overlaying distribution cost and time plots, as well as scatterplots.
- Correlation Type 1 (correlation between repeated activity) caused the most substantial increase in the cost and time of standard deviation of earthwork and pavement structure, which increased by 1,104.35% and 756.055 in cost and time standard deviation of earthwork structure, respectively, and 1167.67% and 401.14% in cost and time of standard deviation of pavement structure,

respectively. On the other hand, correlation Type 2 (correlation between activities) caused the largest increase in the cost and time of standard deviation of bridge structure by 234.56% and 551.42%, respectively.

6 Discussion and Conclusions

This study examined the three types of correlation between the costs, times and cost/time of three main structures of highway projects and modelled their impacts on the distribution of the total cost and the total time of construction of highway structures. The effect of three types of correlations on the total cost and time standard deviation was modeled by Gaussian copula, a strong correlation model which allows correct generation of correlated variables, and the results were simulated by Monte Carlo simulation using actual data of 16.5 kilometers of a new highway constructed in South Africa.

The three structures of the case study have been organized in an arrangement that allows analysis of the effect of the three correlation types in terms of the magnitude of the correlation coefficients, types of the correlation matrix, marginal distributions, and some correlated activity costs. The dominant type of correlation on cost and time of different structures of the highway project was determined by comparing the differences in total cost and time variation of standard deviations. In the earthwork and pavement, correlation Type 1 causes the most substantial increase in total cost and time standard deviation, whereas, in the bridge structure, correlation Type 2 causes the most significant increase in total cost and time standard deviation.

Concerning the magnitude of the correlation coefficients, the results have clearly shown that the more significant the correlation coefficients (ρ=0.8 in the pavement), the larger the increase on the total cost standard deviations (1104.35%), which is aligned with the findings of Moret and Einstein [4] study. However, in the time variable, the combination of the magnitude of the correlation coefficients and arrangement of critical activities on the network activity model of the structure determined the largest increase in the standard deviations of total time, which, in this study, is earthwork because ρ=0.8 and all earthwork activities are located on a critical path (sequential network model without any dummy). The reason for this difference in type of correlation and considerable difference in the variation of standard deviation can be explained with the number of correlated costs and times: correlation Type 1 consists of the correlation between many activities' costs and times, while correlation Type 2 consists of the correlation between few activities' costs and times in bridge structure.

The conclusion is that, as more costs and times are correlated with larger correlation coefficients, the effect on the total cost and time standard deviation becomes more substantial, which finding is supported by the studies of Bakhshi and Touran [2], Moret and Einstein [4]. The results of modelling three types of correlations on the cost and time of a highway project proved that the range of the total cost and total time increase considerably due to correlations. This study has shown that, although estimating the effect of correlations on cost is more straightforward, estimating the effect of correlations on time is more complex. Furthermore, the study proved that the correlation impacts depend on the magnitude of the correlation coefficient, the correlation matrix, and the type of correlation. Also, the impact of the last is tightly connected to the number of activity costs and times that are correlated and the network model of the structure.

By considering the results of modelling correlation on the cost and time distribution of a highway project, it is now possible to determine the effect of such correlation in cost and time prediction models. Modelling correlation is an initial step towards a more comprehensive modelling of uncertainties of construction cost and time. The practical application of the proposed methodology can be extended to any construction project characterized by numerous repetitions of the same activities, particularly to the linear construction projects.

Acknowledgements

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. The opinions expressed, and conclusions arrived at, are those of the authors and are not necessarily to be attributed to the NRF.

References

1. Flyvbjerg B, Holm MS, Buhl S. Underestimating costs in public works projects: Error or lie? Journal of the American planning association. 2002;68(3):279-95.
2. Bakhshi P, Touran A. A Method for Calculating Cost Correlation among Construction Projects in a Portfolio. International Journal of Architecture, Engineering and Construction, 1 (3). 2012:134-41.
3. Touran A. Probabilistic cost estimating with subjective correlations. Journal of Construction Engineering and Management. 1993;119(1):58-71.
4. Moret Y, Einstein H. Modeling correlations in rail line construction. Journal of Construction Engineering and Management. 2012;138(9):1075-84.
5. Purnus A, Bodea C-NJPE. Correlation between time and cost in a quantitative risk analysis of construction projects. 2014;85:436-45.
6. Benjamin JR, Cornell CA. Probability, statistics, and decision for civil engineers: Courier Corporation; 2014.
7. Ökmen Ö, Öztaş A. Construction cost analysis under uncertainty with correlated cost risk analysis model. Construction Management and Economics. 2010;28(2):203-12.
8. Vrac M, Billard L, Diday E, Chédin A. Copula analysis of mixture models. Computational Statistics. 2012;27(3):427-57.