International Journal of Management and Fuzzy Systems
2025, Vol. 11, No. 2, pp. 62-89
https://doi.org/10.11648/j.ijmfs.20251102.12

Research Article

Development of Cost Prediction Model for Improving Cost Estimate at Completion of Building Construction Projects Using MATLAB R2014a Simulink, and Regression Model

Girmay Getawa Ayalew*, Genet Melkamu Ayalew

Department of Construction Technology and Management, School of Civil and Water Resource Engineering, Institute of Technology, Woldia University, Woldia, Ethiopia

Abstract

The success of a project is obtained through the proper management from the beginning to the end of the project. Several studies have been conducted in the project management field further to improve the earned value management methodology to forecast the project cost estimate at completion. However, this study provides to investigate a new research methodology proposed to provide more reliable CEAC by using MATLAB R2014a and Multiple regressions. Thus, the main objective of this study is to focus on the development of a cost prediction model for improving cost estimate at completion of building construction projects using MATLAB R2014a, and regression model. The study is conducted on an EVM data set comprising five real-life projects database gathering between 2021 and 2024. The analysis method was made by using Microsoft excel, MATLAB R2014a, and a statistical package for social science as an analysis tool. The finding of the study revealed that the dependability of EAC on input variable produced by the membership function, and rule viewer is 70% of the estimate at compilation which is quite an acceptable value. It was also found that the regression model shows excellent results in prediction with a coefficient of correlation is 95.70%, 99.90%, 96.10%, 92.40%, and 90.40% for the five projects. Similarly, the coefficient of determination is 91.70%, 99.99%, 92.40%, 85.40%, and 81.70% respectively for the five projects. Finally, it can be recommended that the developed model be conducted in building projects to demonstrate its practicality.

Keywords

Prediction, Cost, Estimate at Completion, MATLAB R2014a, Regression Model

1. Introduction

The success of a project is obtained through the proper management from the beginning to the end of the project. Project management is an art and a science; it demands the proper management of people and knowledge about the various techniques, methods, processes, etc. involved in a project. Every project undergoes the triple constraints of time, money, and quality; the balancing of which results in a satisfactory project output [1]. While a project is being executed monitoring and controlling is very essential. Performance reporting is a part of the project monitoring and control process. The earned value method is a widely accepted performance measurement technique. This methodology is useful

*Corresponding author: girmaygetawa@gmail.com (Girmay Getawa Ayalew)

Received: 21 June 2025; **Accepted:** 7 July 2025; **Published:** 28 July 2025

for reporting the status of the project as well as for predicting future performance based on past performance. Earned value management is one of the potential measurement methods of performance and evaluating the success of a construction project [2].

EVM has been called ''management with the lights on'' because it can help clearly and objectively illuminate where a project is and where it is going as compared to where it was supposed to be and where it was supposed to be going [3]. Forecasting project cost at completion is of great importance to project management success. It is a forward-looking tool to assist Project Managers with the task of making timely and appropriate decisions about cost outcomes of their in-progress projects. For over four decades, Earned Value Management (EVM) has been used to forecast cost at completion.

EVM integrates scope, price, and schedule management underneath an equivalent framework and it provides performance variances and indexes that permit managers to discover over-costs and delays [2]. It is widely applied for measuring and analyzing project actual status against its baseline, and for providing estimates of project cost and duration at completion. In particular, EVM is used to compute Cost Estimate at Completion (CEAC), a top-down estimate of the project's total cost based on the project’s status. Within the EVM framework, several methods exist to compute CEAC, classified as either index-based (IB) or regression-based techniques, statistics-aided, and simulation-based methods [4].

In general, Originating in the 1960s, EVM has become a standard methodology to track a project's status and to estimate the cost at completion based on measured actual performance against scheduled performance. IB methods have an inherent limitation due to their reliance on past information: they assume that the remaining budget is adjusted by a performance index**.** The second concern associated with EACs with statistical tools is to overcome the limitations of index-based methods and to improve the accuracy of EAC forecasting. Some of the advantages of the statistical approach are a proper addressing of the quality of project execution, probabilistic estimations, undefined scope of work, and risk impact [4]. Third simulation techniques are regarded as an alternative approach to both index-based and statistics-aided methods. This approach is of valuable use during the early stages of project execution when conventional methods might fail to accurately predict EAC due to missing cost information and the effects of early risks and uncertainty [4].

However, reported literature reveals that little advancement has been made in the area of improving the reliability of the IB approach via its refinement by regression techniques [5]. Most studies integrating regression concepts into IB approaches concern U.S. defense projects, which are complex with large budgets and long durations [5]. It is accepted that using the EAC technique is useful for reporting the status of the project as well as for predicting future performance based on past performance, even though the studies on the EAC are very limited. [4] studied the cost estimate at completion methods in construction projects using conventional index-based, statistics-aided, and simulation-based cost EAC methods and summarized with characteristics and advantages. [5] conducted an earned schedule-based regression model to improve cost estimates at completion by using the Gompertz growth model. [6] studied integrating exploratory factor analysis and fuzzy AHP models for assessing the factors affecting the performance of building construction projects. [7] studied on prediction of construction cost overrun in Tamil Nadu- a statistical fuzzy approach. [8] conducted on the estimation of cost overrun in construction projects using fuzzy logic. [9] investigated on a fuzzy neural network to estimate at completion costs of construction projects and generate accurate forecasts for the entire future periods using a fuzzy neural network model. [10] conducted a study on developing fuzzy-based earned value analysis model for estimating the performance of construction projects. [11] conducted on regression modeling for prediction of earned value indexes in public building construction projects. However, the previous studies conducted on CEAC do not consider project duration, variation order, addition of works, omission of works, and excess in quantities in cost estimates to generate accurate forecast for the entire future periods using MATLAB R2014a, and Multiple regression.

However, the researchers did not find any study that is the same as this study while searching in international libraries and discreet periodicals such as Scopus, Springer, Taylor France, and others, as this study is almost unique of its kind and deals with CEAC. To fill these gaps, a new research methodology is proposed to provide more reliable CEAC by using MATLAB R2014a and Multiple regressions. Several studies have been conducted in the project management field further to improve the earned value management methodology to forecast the project cost estimate at completion. Thus, the main objective of this study is to focus on the development of a cost prediction model for improving cost estimate at completion of building construction projects using MATLAB R2014a, and regression model. The study provides an important secondary source of data for future researchers and academicians to get knowledge, and skills on the proposed model, and also provides for a reference guide for the building construction projects for evaluation of the project performance.

The remainder of this paper is structured as follows. Section 2 introduces a literature review regarding factors affecting estimate at completion. Section 3 presents the methodology it includes, the method of data collection, fuzzy logic controller, and regression modeling. Section 4 employed a case study to show numerical data analysis for predicting estimates at compilation using fuzzy inference system modeling and regression modeling. Section 5 presents a discussion of the results. Eventually, general conclusions and remarks are then presented in Section 6.

2. Literature Review

Earned value management data can also be used as a forecasting tool in addition to analyzing variances and performance indices for trends to determine problems in the execution of a project [12]. During the execution of a project, forecasting has two objectives: to provide a forecast for the outcome of the project based on current status and trends; and second, to highlight trends or potential budget deviations that require management control. When forecasting, one has to take into account the time phase of the project. At the beginning of a project, uncertainty about the outcome is higher than at the end of the project.

This uncertainty results in a range of possible final projections. With earned value data, however, early predictions have proven to be reliable [12]. The Estimate at Completion (EAC) is an important tool used in forecasting because it indicates where the project cost is heading [12]. Many studies have been carried out to introduce various new or improved EAC techniques. According to [4], the EAC forecasting method would be divided into three main categories, namely: traditional index-based, statistics-aided, and simulation-based methods. [13] constructed an evolutionary EAC model to generate project cost estimates that proved significantly more reliable than estimates achievable using currently prevailing formulae. [1] compared the performance of the Support Vector Regression model with the standard EVM one showing the superiority of the former method. [14] introduced two additional multivariate control metrics that allow the dynamic monitoring of the project, suggesting the need for additional metrics to describe the EVM/ES management system correctly. An initial attempt to adjust the EAC was made by [15], the study compared nine types of regression models and identified the one that best suited a cost prediction model for the reconstruction project. [10] stated that developing a fuzzy-based earned value analysis model for estimating the performance of projects to make estimations get closer and reduce the errors in calculations through the fuzzy linguistic terms. [11] stated on regression modeling for the prediction of earned value indexes in public building construction projects. [13] provided a different approach to the EAC problem, developing a model to predict the specific AC and EV by proposing a simple formula for data transformation.

2.1. Factors Affecting Estimate at Completion

Estimate at Completion (EAC) was known as the subject factor and each project is used as the basic unit of oversight [16]. Modeling of the factors affecting Estimate at Completion (EAC) has become a crucial case facing the construction stakeholders for a long time, to put more effort and attention into the control of the construction projects.

Estimate at Completion (EAC) is a largely used technique that has been showing itself as an objective and valuable tool to accurately compute how much the project is going to be in the finish associated with the key data points of Earned Value Management (EVM). Estimate at Completion (EAC) is the projected final cost required to finish the complete work and is based on a statistical prediction using the performance indices.

$$EAC = BAC / CPI \tag{1}$$

Variance at Completion (VAC) is the variance in the total budget at the end of the project. It is the difference between what the project was originally planned to cost, versus what it is now estimated to cost.

$$VAC = BAC – EAC \tag{2}$$

The key data points in earned value management that affect Estimate at Completion (EAC) were discussed as follows [17].

2.2. Planned Value (PV)

Planned Value (PV) describes how far along project work is supposed to be at any given point in the project schedule. It is also known as Budgeted Cost of Work Performed (BCWP).

2.3. Earned Value (EV)

Earned Value (EV) is a snapshot of work progress at a given point in time. Also known as the Budgeted Cost of Work Performed (BCWP), it reflects the amount of work that has been accomplished to date (or in a given period), expressed as the planned value for that work.

2.4. Actual Cost (AV)

Actual Cost (AC), also known as the Actual Cost of Work Performed (ACWP), is an indication of the level of resources that have been expended to achieve the actual work performed to date (or in a given period).

1) Schedule Variance (SV)

Schedule Variance (SV) is the difference between the planned value of the work scheduled and the value of the work accomplished for the same time phase. It displays objectively how much the project is ahead or behind schedule.

$$SV= EV – PV \tag{3}$$

SV > 0, ahead of schedule

SV < 0, behind of schedule

2) Cost Variance (CV)

Cost Variance (CV) is defined as the difference between the values of the work accomplished and the actual cost incurred to perform the work; utilizing this parameter, the percentage of cost overrun or under run can be calculated.

$$CV= EV – AC \quad (4)$$

CV > 0, under budget
CV < 0, over budget
3) Schedule Performance Index (SPI)
Schedule Performance Index (SPI) is an index showing the efficiency of the time utilized on the project.

$$SPI = EV/PV \quad (5)$$

SPI > 1, efficiency in utilizing the time allocated to the project is good
SPI < 1, efficiency in utilizing the time allocated to the project is poor
4) Cost Performance Index (CPI)
The Cost Performance Index (CPI) is an index showing the efficiency of the utilization of the resources allocated to the project.

$$CPI = EV/AC \quad (6)$$

CPI > 1, efficiency in utilizing the resources allocated to the project is good
CPI < 1, efficiency in utilizing the resources allocated to the project is poor

2.5. Budget at Completion (BAC)

Budget at Completion (BAC) is the cost of the total estimated work in the plan, located at the end of the PV curve.

Estimate to Complete (ETC) is the estimated cost required to finish all the remaining work, calculated when the past estimating assumptions become invalid and a need for revised estimates.

$$ETC = (BAC – EV) / CPI \quad (7)$$

2.6. Project Duration

Project duration is the total time required to complete a project from start to finish. Project duration is a critical aspect of project planning and scheduling, referring to the total time needed to complete a project from its inception to its conclusion [18]. According to [19] project duration is defined as the measure of time that is necessary to complete a project. Project duration is calculated with the help of a project's timeline that may include parallel and successive activities, so the whole amount of time being covered by a project's timeline from the beginning to its end is the duration of this project, without relation to a number and durations of all encased activities.

2.7. Variation Order

A variation order is a written agreement to modify, add to, or otherwise alter the work from that outlined in the contract [20]. In general, variation orders involve changes in scope, time, and material for additional work and extra costs for additional working hours. Due to the dynamic and complex structure of the construction industry changes in the construction projects are inevitable and variation orders are utilized in all types of projects.

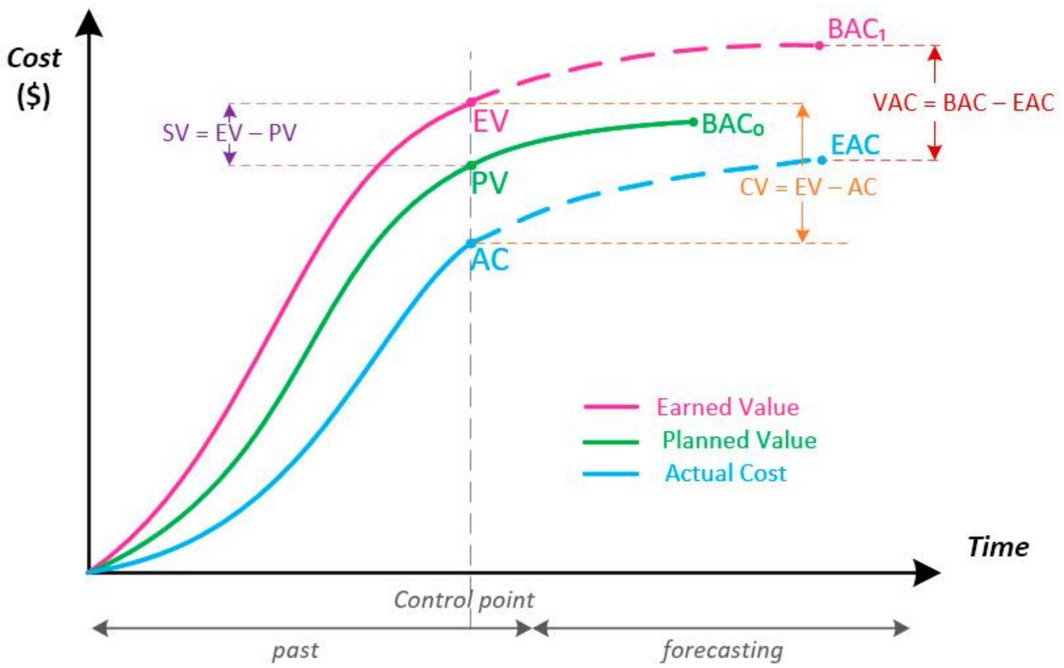


Figure 1. Standard Earned Value Analysis Graph.

3. Methodology

The brief methodology that was implemented in this research paper on the development of a cost prediction model for improving cost estimate at completion of building construction projects using an Interactive Fuzzy Logic by using MATLAB R2014a Software, and a regression model. Interactive Fuzzy logic is becoming an extremely suitable and satisfactory tool for the development of real-time applicable expert systems in the construction industry providing appropriate reasoning to complex computational problems by developing user-friendly interfaces. Multiple linear regression analysis is performed to evaluate the number of regressors, the priority of the candidate EVM variables in the regression model, and to assess the diagnostics of the model fit. This study can employ a case study research approach made on some projects selected based on the standard data set, and it collects the schedule, cost, and variation-related information from the public building construction projects for predicting the Estimate at Completion (EAC) of the building construction projects.

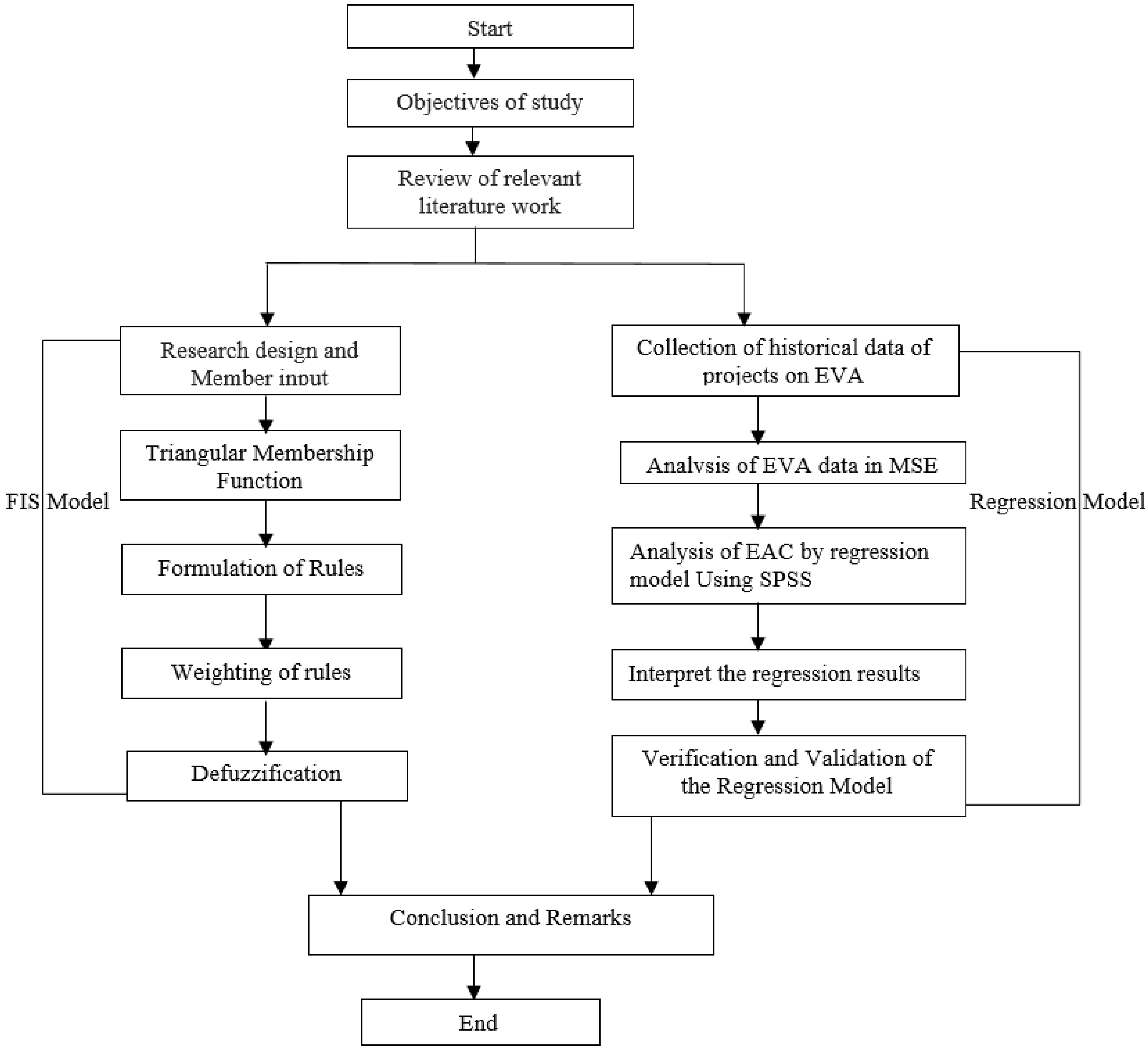


Figure 2. *The steps, and procedures conducted by the applied methodologies.*

3.1. Method of Data Collection

Data collection was carried out by conducting a questionnaire survey to identify the most significant factors affecting the Estimate at Completion from the perspective of project managers, consultants, and contractors. Moreover, the method of data collection used in this study was based on the historical data of five building construction projects which were done between 2021 and 2024 by project managers, consultants, and contractors. In this study, the collected data has been analyzed by using SPSS (Statistical package for social sciences) using the historical data taken from building projects. The key data points in earned value management that affect Estimate at Completion have been analyzed through the fuzzy logic toolbox of MATLAB R2014a program software, and the regression model. MATLAB R2014a software and Statistical package for social sciences (SPSS) were used as the basic tools for all sorts of calculations.

3.2. Identification of Study Variables

The two types of variables considered in developing the cost prediction model for improving cost estimate after building construction projects are dependent variables and Independent variables.

Independent variables: The variable that is manipulated or controlled by the researcher. It is the variable that is hypothesized to affect the dependent variable; these variables are also called predictor or explanatory variables used to predict or explain the behavior of the dependent variable [21]. In this study, the actual cost, earned value, planned value, budget at completion, duration of the project, and variation orders were identified as the independent variables.

Dependent variables: is the variable that is being measured or observed in the study. The dependent variable is the variable that is expected to change as a result of the independent variable manipulation [22]. In this study, the dependent variable is Estimate at Completion (EAC).

3.3. Method of Data Analysis

The method of data analysis in this study was briefly discussed in the following sub-section.

3.3.1. Relative Importance Index

The Relative Importance Index (RII) is the principal statistical tool that is used to determine the respondents' perception a five-point Likert scale of 1-5 was adopted to assess the degree of agreement of each factor where 1 represents "very low", 2 - low, 3- medium, 4-high and 5-very high.

The important index is evaluated using the formula [23]:

$$\text{Relative Importance Index (RII)} = \sum \frac{\text{Wixi}}{\text{AN}} \quad (8)$$

Where RII= is the relative importance index, RII ranges from zero to one.

W = is the weight given to each factor by the respondents and ranges from 1 to 5

x = is the response given for each weight

A = the highest weight = 5

N = the total number of respondents.

According to [24] the range of a set of data is the difference between the largest and smallest values and the class width of the relative importance index values can be determined by dividing the range by a five-point scale which is five to classify the factors in to very high, high, medium, low, and very low.

$$\text{Range= maximum RII – minimum RII} \quad (9)$$

$$\text{Width= (maximum RII – minimum RII)/5} \quad (10)$$

3.3.2. Fuzzy Logic Controller

The concept of fuzzy logic was first introduced in 1965 by Zadeh. The term "fuzzy" refers to the fact that the logic involved can deal with concepts that cannot be expressed as "true" or "false" but rather as "partially true". A fuzzy logic system (FLS) can be defined as the nonlinear mapping of an input data set to a scalar output data. A FLS consists of four main parts: fuzzifier, rules, inference engine, and defuzzifier. This component of a FLS is shown in Figure 4. Firstly, a crisp set of input data is gathered and converted to a fuzzy set using fuzzy linguistic variables, fuzzy linguistic terms, and membership functions. This step is known as fuzzification. Afterward, an inference is made based on a set of rules. Lastly, the resulting fuzzy output is mapped to a crisp output using the membership functions, in the defuzzification.

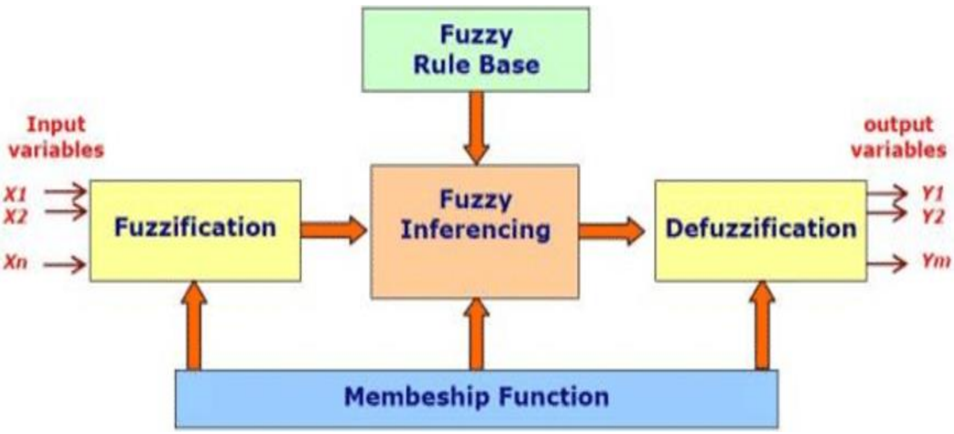


Figure 3. *A Fuzzy Logic System.*

a. Fuzzy Sets and Membership Functions

Membership functions are used in the fuzzification and defuzzification steps of a FLS, to map the non-fuzzy input values to fuzzy linguistic terms and vice versa. A membership function is used to quantify a linguistic term. In a fuzzy set, each object has its membership value, which determines the degree to which the object belongs to a fuzzy set [25]. Membership values range between 0 and 1. Fuzzy set proposed a theory from ordinary crisp sets with a membership value of either 0 or 1. In fuzzy sets, linguistic values are quantified by the implementation of fuzzy rules. A membership function is a curve in which input space is sometimes referred to as the universe of discourse, defining how each point in the input space is mapped to a membership value between 0 and 1. Triangular, angular, and trapezoidal shape membership functions are different types of membership functions [7].

b. Fuzzy Rules

In an FLS, a rule base is constructed to control the output variable. A fuzzy rule is a simple IF-THEN rule with a condition and a conclusion [7].

c. Fuzzy Set Operations

The evaluations of the fuzzy rules and the combination of the results of the individual rules are performed using fuzzy set operations. The operations on fuzzy sets are different than the operations on non-fuzzy sets. The classical set has a sharp boundary, which means that a member either belongs to that set or does not. Also, this classical set can be mapped to a function with two elements, 0 or 1. In other words, this “fully belonging to” can be mapped as a member of set A with a degree of 1, and “not belonging to” can be mapped as a member of set A with a degree of 0 [26]. This mapping is similar to a black-and-white binary categorization. Compared with a classical set, a fuzzy set allows members to have a smooth boundary. In other words, a fuzzy set allows a member to belong to a set to some partial degree [8].

The basic fuzzy set operations also include intersection, union, and complement, and those operations are defined as [7].

$$\text{Union: } A \cup B = \mu_A(x) \cup \mu_B(x) = \max(\mu_A(x), \mu_B(x) \tag{11}$$

$$\text{Intersection: } A \cap B = \mu_A(x) \cap \mu_B(x) = \min(\mu_A(x), \mu_B(x) \tag{12}$$

$$\text{Complement: } A^c = F \backslash A \tag{13}$$

Where A and B are two fuzzy sets and x is an element in the universe of discourse, X. According to the definition of fuzzy set operations, a fuzzy set union operation is exactly equivalent to selecting the maximum member from those members in the sets, and the intersection operation is selecting the minimum member from the sets.

Steps Involved in the Fuzzy Logic Analysis

To develop the model, the following steps are performed on the fuzzy logic tool box of MATLAB R2014a.

a. Construct a six-input and one-output system in the FIS editor. The identified key input data points in earned value management and “Estimate at Completion (EAC)” are entered as input members and output members respectively.

b. Membership functions associated with all of the input and output variables are defined in the membership function editor.

c. To perform fuzzy inference, rules that connect input variables to output variables are defined. For the present model, 21 rules are constructed in the form of IF-THEN.

3.3.3. Predicting Estimate at Completion by Using Regression Analysis

The main purpose of this analysis is to know to what extent is the Estimate at Completion influenced by the six independent variables and what are measures should be taken based on the results obtained using SPSS - Statistical Package for Social Sciences [27]. Statistics and data analysis need to establish a relationship between the various parameters in a data set for prediction and analysis. It involves fitting the right model concerning the given data set and then using that model to make further predictions [28].

a. Simple Linear Regression Model

The modeling between the dependent and one independent variable. When there is only one independent variable in the linear regression model, the model is generally termed a simple linear regression model. When there is more than one independent variable in the model, then the linear model is termed the multiple linear regression model.

Consider a simple linear regression model.

$$y = \beta 0 + \beta 1 X \ \varepsilon \tag{14}$$

b. Multiple Regression Analysis

Multiple Regression Analysis refers to a set of techniques for studying the straight-line relationships among two or more variables. Multiple regression estimates the β’s in the equation [13].

$$y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \ldots + \beta_p x_{pj} + \varepsilon \tag{15}$$

The X's are the independent variables (IVs). Y is the dependent variable. The subscript j represents the observation (row) number. The β's are the unknown regression coefficients. Their estimates are represented by b’s. Each β represents the original unknown (population) parameter, while b is an estimate of this β. The ε_j is the error (residual) of observation j.

4. Case Study: Numerical Data Analysis

4.1. Demographic Analysis

The characteristics of the respondents participating in the survey are summarized in Table 1. As shown in Table 1, 26.67% were project managers, 46.67% were contractors, and 26.67% were consultants. The educational level of the respondents indicates that 83.33% had an MSc 16.67% of the respondents were PhD holders. The results of the study presented in the table show that 73.33% of the respondents were male and 26.67% were female. Hence, the respondents were not gender biased, and the sampling technique ensured the inclusion of all members of the population being sampled for the study. Regarding the working experience, 33.33% of the respondents have between 5 and 10 years of working experience, 40.00% of the respondents have between 11 and 15 years of working experience, and 26.67% of the respondents have greater than 15 years of working experience.

Table 1. *The expert panel's demographics.*

Parameter	Frequency	Percentage (%)	Parameter	Frequency	Percentage (%)
Type of organization			Gender		
Project managers	8	26.67%	Male	22	73.33%
Contractors	14	46.67%	Female	8	26.67%
Consultants	8	26.67%	Work experience of respondents		
Educational level			6-10 years	10	33.33%
MSc	25	83.33%	11-15 years	12	40.00%
PhD	5	16.67%	$\geq$15 years	8	26.67%

4.2. Reliability and Validity of Questionnaire

4.2.1. Reliability of Questionnaire

Reliability tests were conducted at the beginning of the analysis of results to check the reliability, and internal consistency of data before they were analyzed by using Cronbach's alpha. The reliability coefficient normally ranges between 0 and 1. According to [29] when Cronbach's Alpha is more than 0.7 there is high internal consistency for the data set. The formula that determines alpha is fairly simple and makes use of the items (variables), k, in the scale and the average of the inter-item correlations, r. The conceptual formula of Cronbach's Alpha is defined by [30].

$$\alpha = \frac{Kr}{1+(K-1)r} \tag{16}$$

The value of Cronbach's alpha is 0.828 for the entire questionnaire as shown in Table 2 which indicates excellent reliability of the entire questionnaire.

Table 2. *Reliability Statistics.*

Cronbach's Alpha	N of Items
.828	6

4.2.2. Correlation Test

Pearson correlation test has been performed to examine the relationship between the factors having an impact on the estimate at compilation. The purpose of this test is to examine and evaluate the relationships involving the factors contributing to the estimate at compilation. As it was observed in Table 3 the variables were highly correlated with a high correlation coefficient value and significance value.

Table 3. *Pearson correlation relationship between variables.*

Variables		Planned Value	Earned Value	Actual Cost	Budget at Completion	Project duration	Variation order
Planned Value	Pearson Correlation	1	.400*	.466**	.554**	.625**	.377*
	Sig. (2-tailed)		.028	.009	.001	.000	.040
	N	30	30	30	30	30	30

Variables		Planned Value	Earned Value	Actual Cost	Budget at Completion	Project duration	Variation order
	Pearson Correlation	.400*	1	.655**	.243	.377*	.189
Earned Value	Sig. (2-tailed)	.028		.000	.196	.040	.316
	N	30	30	30	30	30	30
	Pearson Correlation	.466**	.655**	1	.383*	.425*	.387*
Actual Cost	Sig. (2-tailed)	.009	.000		.037	.019	.035
	N	30	30	30	30	30	30
	Pearson Correlation	.554**	.243	.383*	1	.442*	.804**
Budget at Completion	Sig. (2-tailed)	.001	.196	.037		.014	.000
	N	30	30	30	30	30	30
	Pearson Correlation	.625**	.377*	.425*	.442*	1	.406*
Project duration	Sig. (2-tailed)	.000	.040	.019	.014		.026
	N	30	30	30	30	30	30
	Pearson Correlation	.377*	.189	.387*	.804**	.406*	1
Variation order	Sig. (2-tailed)	.040	.316	.035	.000	.026	
	N	30	30	30	30	30	30

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

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

4.3. Results of Factors Affecting Estimate at Compilation

The findings of the study obtained from the analysis of the questionnaire show that the factors affecting estimate at compilation were identified and ranked on the perception of the respondents were presented in Table 4 below.

Table 4. *Preliminary ranking of the factors affecting estimate at compilation.*

S/N	Factors Affecting Estimate at Compilation	RII	Ranking
1	Planned Value	0.620	6
2	Earned Value	0.700	4
3	Actual Cost	0.773	1
4	Budget at Completion	0.747	2
5	Project duration	0.727	3
6	Variation order	0.680	5

The findings obtained from the questionnaire analysis presented in the above Table 5 was further can be classified in to very high, high, medium, low, and very low according to the respondent's opinions as follows.

Table 5. *Categories of factors affecting estimate at compilation.*

Scale	Class boundaries	Respondents Opinions
1	$0.620 \leq RII \leq 0.651$	Very low
2	$0.651 < RII \leq 0.681$	Low
3	$0.681 < RII \leq 0.712$	Medium
4	$0.712 < RII \leq 0.743$	High
5	$0.743 < RII \leq 0.773$	Very high

The finding of the study shown in Table 4 implied that the respondents classified the factors affecting estimate at compilation is planned value under very low factors, variation order under low factors, earned value under medium factors, project duration under high factors, and actual cost, and budget at completion were identified as very high factors.

4.3.1. Predicting Estimate at Compilation Using Fuzzy Inference System Modeling

The cost prediction model for improving cost estimate after building construction projects was studied in this work, the fuzzy controller involved in analyzing and predicting the estimate at compilation based on the six factors. The six input and one output fuzzy-based estimates at completion are discussed here. Table 6 shows the linguistic variables used in the model and their membership function. The overall factors in estimate at compilation and its corresponding weight factors are tabulated in Table 6. The six factors are used as inputs and estimated at compilation as the output of the fuzzy controller. In the fuzzy logic controller, all membership functions are considered as triangular membership functions with five segments. The fuzzy subsets are Very Low, Low, Medium, High, and Very High. The Mamdani inference system is used in the fuzzy controller. Figure 5 shows the Flow diagram for the development of an estimate at compilation model in a fuzzy inference system. The fuzzy inference system is shown in Figure 5 and the membership function editor for estimate at compilation is shown in Figure 6.

Table 6. *Linguistic variables used in the model and their membership function.*

Variables	Range	MFs	No of MFs	Name of the parameters
Planned Value	[0 -100]	Trimf	5	1. very low 2. low 3. medium 4. high 5. very high
Earned Value	[0 -100]	Trimf	5	1. very low 2. low 3. medium 4. high 5. very high
Actual Cost	[0 -100]	Trimf	5	1. very low 2. low 3. medium 4. high 5. very high
Budget at Completion	[0 -100]	Trimf	5	1. very low 2. low 3. medium 4. high 5. very high
Project duration	[0 -100]	Trimf	5	1. very low 2. low 3. medium 4. high 5. very high
Variation order	[0 -100]	Trimf	5	1. very low 2. low 3. medium 4. high 5. very high

To develop the model, the following rules are performed on the fuzzy logic tool box of MATLAB R2014a. To perform fuzzy inference, rules that connect input variables to output variables are defined. In this study, 21 rules are constructed in the form of IF-THEN. Five of them are presented in Table 7 below.

Table 7. *Sample fuzzy rules for the cost assessment model and rules weight.*

S.No	Rules	Rule weight
1	If the probability of a planned value is very low the estimate at compilation is very low	0.620
2	If the probability of an earned value is very low the estimate at compilation is very low	0.700
3	If the probability of actual cost is very low the estimate at compilation is very low	0.773
4	If the probability of budget at completion is very low the estimate at compilation is very low	0.747
5	If the probability of project duration is very low the estimate at compilation is very low	0.727
6	If the probability of variation order is very low the estimate at compilation is very low	0.680

The estimate at completion model was developed considering the six factors using the Fuzzy Logic Designer tool that is available in MATLAB R2014a. The analysis steps for the development of the model are discussed as follows. The first step of the fuzzy inference system is defining the input and output variables for the model development. In the FIS editor of MATLAB R2014a, a six-input and one-output system was developed in Figure 5 as follows.

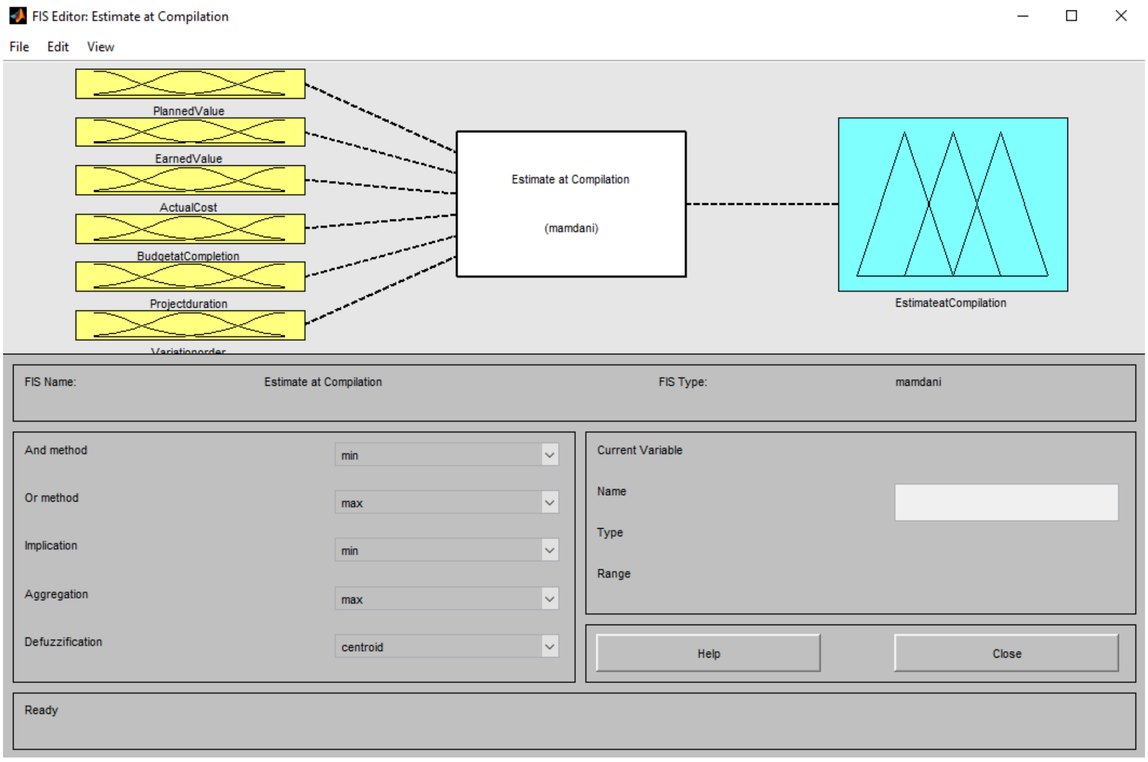


Figure 4. Fuzzy Inference System.

In the second step, the input variables were fuzzified with the help of membership functions. Membership functions for each input and output variable were aligned in the Triangular form of function wherein each input variable was given 5 linguistic divisions very low, low, medium, high, and very high similarly the output variable consisted of 5 linguistic divisions very low, low, medium, high, and very high. Ranges of these divisions were defined according to the findings from the literature review and suggestions from experienced personnel in the industry.

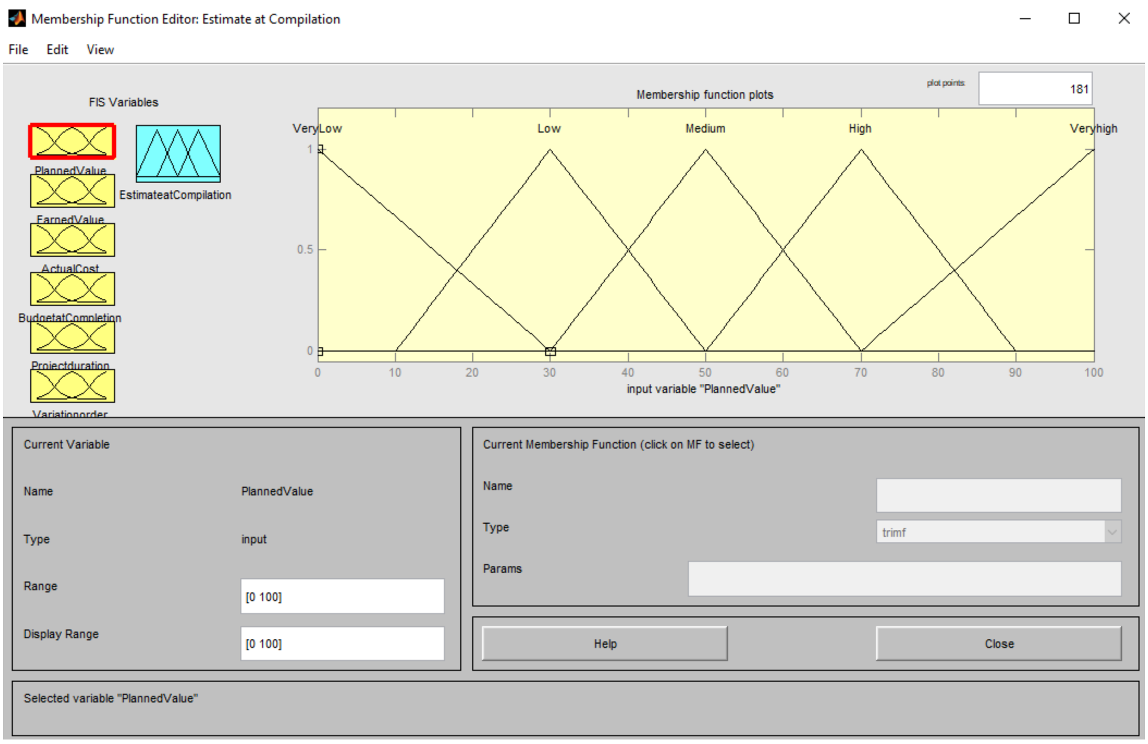


Figure 5. Membership Function Editor.

In the third step, fuzzy rules that connect input variables to output variables were defined to perform the inference. In this study, 21 rules have been constructed in the form of IF- then. The relative percentage based on the Relative Importance Index (RII) of the variables that affect the estimate at compilation was found and assigned as weightage to the fuzzy rules to develop the assessment model. Since the obtained RI values for the six factors were found to be less distinct, the ranges of input and output linguistic variables were defined in the light of the expert's opinion. This method has been useful in achieving more accurate fuzzy rule weights resulting in each IF-Then rule to differ indicating the relative importance of fuzzy rules in the model making.

Figure 6. *Rule View.*

Analysis and Dependability of variables

Assigning of membership function range and making the rule determine the dependability of the result produced by the rule viewer. In this study planned value, earned value, actual cost, budget at completion, project duration, and variation order were used as input variables, and estimate at compilation was used as output in terms of membership function value [0, 100] can be obtained. The input values [76.97, 81.46, 81.46, 76.97, 72.47, 70.22] produce 70% of the estimate at compilation which is quite an acceptable value.

4.3.2. Predicting Estimate at Compilation Using Regression Model

Input variables have the most important influence on the model performance as the significant step in evolving regression models. In this study, estimates at compilation model were developed based on the historical data obtained from five building construction projects. The development of mathematical formulation has been established in this section to predict an estimate at compilation of building projects by using SPSS version 23 as an analysis tool to build up the models for each of the building projects as follows:

Project One: Estimate at compilation Regression Model

The correlation coefficient (R) and coefficient of determination (R^2) for this project is 95.70%, and 91.70% respectively which indicates an excellent level of prediction. The results of the correlation coefficient and coefficient of determination indicate a positive relationship and a high correlation between the dependent variable and the independent variable. As it was observed in Table 8 the value of R^2 is 0.917, this means the independent variables involved in the model explained 91.70% of the variability of the dependent variable, and 8.3% (100%-91.70%) of the variation is caused by factors other than the predictors included in this model. Adjusted R^2 is another important factor that provides a more realistic estimate of the model accuracy than R^2. The value of .915 indicates that 91.5% variance in the dependent variable that is explained by the independent variables. The standard error of the estimate is a statistical measure used to assess the accuracy of predictions made by a regression model. The standard error of a model fit is a measure of the precision of the model. It shows how wrong one could use the regression model to make predictions or to estimate the dependent variable or variable of interest. In this estimate of the estimate at compilation regression model, the model will be wrong by .02576 which is a small amount. The regression statistics of the estimate at compilation regression model for project one (EAC-Model) are shown in Table 8.

Table 8. Entered/Removed Variables for the EAC Regression model.

Variables Entered/Removed			
Model	**Variables Entered**	**Variables Removed**	**Method**
1	VO, PD, BAC, EV, AC, PV[b]		Enter

a. Dependent Variable: EAC

b. All requested variables entered.

Table 9. Model Summary.

Model Summary[b]									
Model	**R**	**R Square**	**Adjusted R Square**	**Std. Error of the Estimate**	**Change Statistics**				
					R Square Change	**F Change**	**df1**	**df2**	**Sig. F Change**
1	.957a	.917	.915	.02576	.932	2554.116	6	15	.000

a. Predictors: (Constant), VO, PD, BAC, EV, AC, PV

b. Dependent Variable: EAC

Table 10. Statistical significance of the model.

ANOVA[a]						
Model		**Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**
1	Regression	101731834889365120.000	6	16955305814894186.000	2554.116	.000[b]
	Residual	99576365881444.450	15	6638424392096.297		
	Total	101831411255246560.000	21			

a. Dependent Variable: EAC

b. Predictors: (Constant), VO, PD, BAC, EV, AC, PV

Table 10 shows that the F-ratio tests whether the overall regression model is a good fit for the data and also shows that the independent variables statistically significant prediction of the dependent variable, $F(6, 15) = 2554.116$, and $p (0.000) < 0.05$, This indicates that the regression model is a good fit of the data.

Statistical significance of independent variables indicates that changes in the independent variables correlate with shifts in the dependent variable. Table 11 shows that the value of ($p < 0.05$) independent variables in this model is statistically significant and already appears in the model except for variation order, which indicated that it was making a significant contribution to the prediction of EAC. The standard error of the coefficients in regression output is also required to be as small as possible. It reflects how wrong could be while estimating its value. In this model relative to the coefficient of the variables, their standard error is small. The prediction formula was generated by multiple regression analysis with the information provided in the coefficients under Table 11.

Table 11. *Statistical significance of the independent variables.*

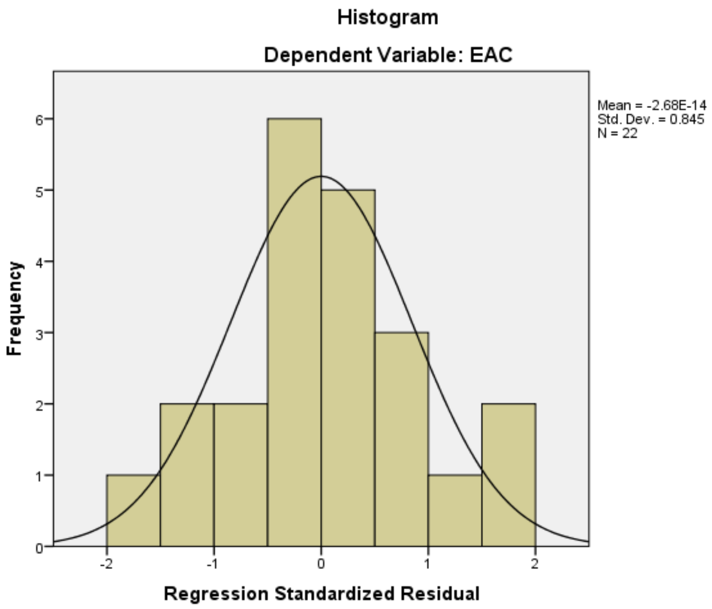
Coefficients[a]								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-38255464.993	14.340		-1.227	.000	-104721083.544	28210153.557
	PV	-.123	.241	-.020	-.511	.000	-.636	.390
	AC	.060	.074	.019	.803	.000	-.099	.218
	EV	-2.359	.159	-.255	-14.807	.030	-2.698	-2.019
	BAC	2.941	.100	1.171	29.508	.000	2.729	3.154
	PD	1439.168	.477	.003	.302	.029	-8711.810	11590.146
	VO	204.084	468.804	.005	.435	.670	-795.148	1203.315

a. Dependent Variable: EAC

Statistics of regression for model EAC are generated in Table 11, which can be formulated as Equation (17) below:

$$\text{Predicted EAC}=-38255464.993-0.123PV+0.060AC-2.359EV+2.941BAC+1439.168PD \quad (17)$$

Standardized coefficients are called beta weights that rank the predictor variables based on their contribution (irrespective of sign) in explaining the outcome variable. In this model, BAC is the highest contributing (1.171) predictor to explain EAC, and AC (0.019) is the next highest predictor variable.



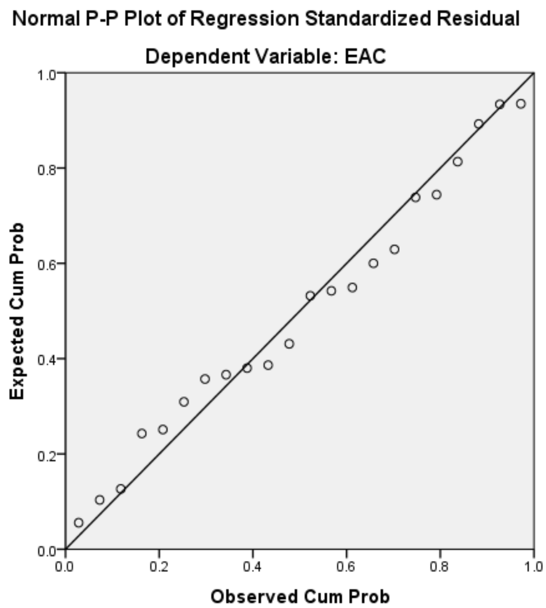


Figure 7. *Histogram of Residuals for the EAC regression model and Plot of observed EAC against expected EAC for P1.*

Project Two: Estimate at compilation Regression Model

The correlation coefficient (R) and coefficient of determination (R^2) for project two is 0.999, and 0.999 respectively which indicates an excellent level of prediction. The results of the correlation coefficient and coefficient of determination indicate a positive relationship and a high correlation between the dependent and the independent variable. As it was observed in Table 12 the value of R^2 is 0.999, this means the independent variables involved in the model explained 99.9% of the variability of the dependent variable, and 0.1% (100%-99.9%) of the variation is caused by factors other than the predictors included in this model. Adjusted R^2 is another important factor that provides a more realistic estimate of the model accuracy than R^2. The value of 0.998 indicates that 99.8% variance in the dependent variable which is explained by the independent variables. The standard error of the estimate is a statistical measure used to assess the accuracy of predictions made by a regression model. The standard error of a model fit is a measure of the precision of the model. It shows how wrong one could use the regression model to make predictions or to estimate the dependent variable or variable of interest. In this estimate of the estimate at compilation regression model, the model will be wrong by 0.067 which is a small amount. The regression statistics of the estimate at compilation regression model for project two (EAC-Model) are shown in Table 12.

Table 12. *Entered/Removed Variables for the EAC Regression model.*

Variables Entered/Removed[a]			
Model	**Variables Entered**	**Variables Removed**	**Method**
1	VO, PD, PV, EV, AC, BAC[b]	.	Enter

a. Dependent Variable: EAC

b. All requested variables entered.

Table 14. *Statistical significance of the model.*

ANOVA[a]						
Model		**Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**
1	Regression	450952748145065470.000	6	75158791357510912.000	2232.012	.000[b]
	Residual	505096729983605.250	15	33673115332240.350		
	Total	451457844875049090.000	21			

a. Dependent Variable: EAC

b. Predictors: (Constant), VO, PD, PV, EV, AC, BAC

Table 14 shows that the F-ratio tests whether the overall regression model is a good fit for the data and also shows that the independent variables statistically significant prediction of the dependent variable, $F(6, 15) = 2232.012$, and $p (0.000) < 0.05$, This indicates that the regression model is a good fit of the data.

Table 15. *Statistical significance of the independent variables.*

Coefficients[a]								
Model		**Unstandardized Coefficients**		**Standardized Coefficients**	**t**	**Sig.**	**95.0% Confidence Interval for B**	
		B	**Std. Error**	**Beta**			**Lower Bound**	**Upper Bound**
1	(Constant)	-33981367.754	0.080		-.854	.007	50843756.759	50843756.759
	PV	-.075	.027	-.029	-.330	.046	.409	.409
	AC	1.046	.074	.395	14.195	.000	1.203	1.203
	EV	-1.532	.136	-.115	-11.286	.000	-1.243	-1.243
	BAC	1.618	.195	.662	8.275	.000	2.034	2.034
	PD	5378.763	10124.842	.005	.531	.603	26959.352	26959.352
	VO	-67.485	1419.634	.000	-.048	.963	2958.393	2958.393

a. Dependent Variable: EAC

Statistical significance of independent variables indicates that changes in the independent variables correlate with shifts in the dependent variable. Table 15 shows that the value of ($p < 0.05$) independent variables in this model is statistically significant and already appears in the model except for project duration, and variation order, which indicated that it was making a significant contribution to the prediction of EAC. The standard error of the coefficients in regression output is also required to be as small as possible. It reflects how wrong could be while estimating its value. In this model relative to the coefficient of the variables, their standard error is small. The prediction formula was generated by multiple regression analysis with the information provided in the coefficients under Table 15.

Statistics of regression for model EAC are generated in Table 15, which can be formulated as Equation (18) below:

$$\text{Predicted EAC} = -33981367.754 - 0.075PV + 1.046AC - 1.532EV + 1.618BAC \quad (18)$$

Standardized coefficients are called beta weights that rank the predictor variables based on their contribution (irrespective of sign) in explaining the outcome variable. In this model, BAC, and AC are the highest contributing predictor variables to explain EAC with Standardized coefficient values of 0.662 and 0.395 respectively.

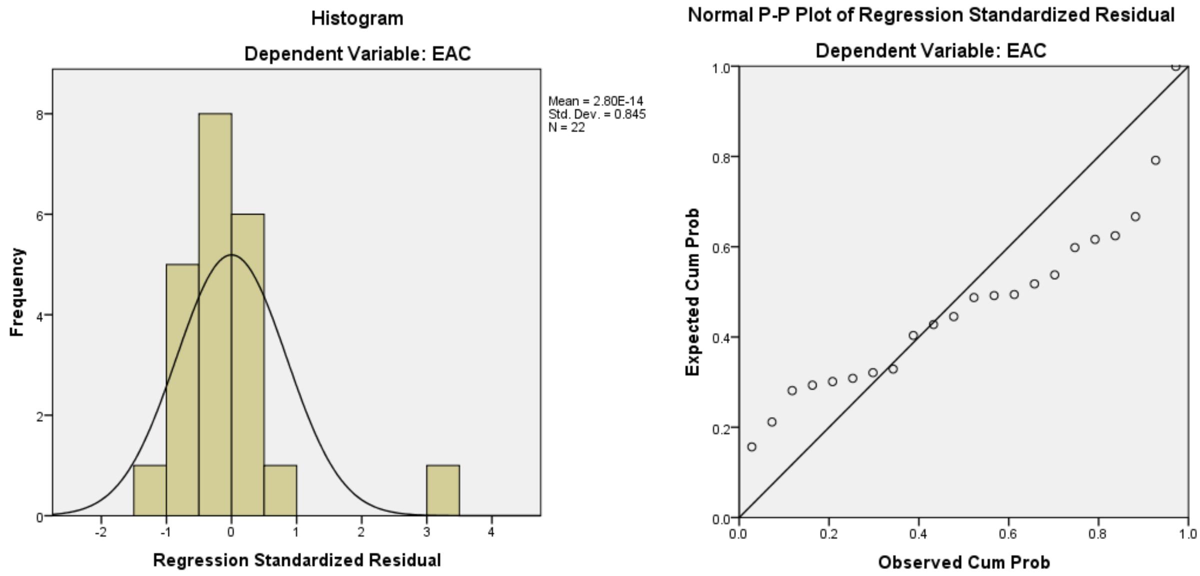


Figure 8. *Histogram of Residuals for the EAC regression model and Plot of observed EAC against expected EAC for P2.*

Project Three: Estimate at compilation Regression Model

Table 16. *Entered/Removed Variables for the EAC Regression model.*

Variables Entered/Removed[a]			
Model	**Variables Entered**	**Variables Removed**	**Method**
1	VO, AC, BAC, PD, EV, PV[b]	.	Enter

a. Dependent Variable: EAC

b. All requested variables entered.

The correlation coefficient (R) and coefficient of determination (R^2) for project three is 0.961, and 0.924 respectively which indicates an excellent level of prediction. The results of the correlation coefficient and coefficient of determination indicate a positive relationship and a high correlation between the dependent and the independent variable. As it was observed in Table 16 the value of R^2 is 0.924, this means the independent variables involved in the model explained 92.4% of the variability of the dependent variable, and 7.6% (100%-92.4%) of the variation is caused by factors other than the predictors included in this model. Adjusted R^2 is another important factor that provides a more realistic estimate of the model accuracy than R^2. The value of adjusted R2 is 0.894 indicating that 89.4% variance in the dependent variable is explained by the independent variables. The standard error of the estimate is a statistical measure used to assess the accuracy of predictions made by a regression model. The standard error of a model fit is a measure of the precision of the model. It shows how wrong one could use the regression model to make predictions or to estimate the dependent variable or variable of interest. In this estimate of the estimate at compilation regression model, the model will be wrong by 0.051 which is a small amount. The regression statistics of the estimate at compilation regression model for project three (EAC-Model) are shown in Table 16.

Table 17. *Model Summary.*

Model Summary[b]									
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.961[a]	.924	.894	.051281	.924	30.416	6	15	.000

a. Predictors: (Constant), VO, AC, BAC, PD, EV, PV

b. Dependent Variable: EAC

Table 18. *Statistical significance of the model.*

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	346415563880428610.000	6	57735927313404768.000	30.416	.000[b]
	Residual	28473150815420224.000	15	1898210054361348.200		
	Total	374888714695848830.000	21			

a. Dependent Variable: EAC

b. Predictors: (Constant), VO, AC, BAC, PD, EV, PV

Table 18 shows that the F-ratio tests whether the overall regression model is a good fit for the data and also shows that the independent variables statistically significant prediction of the dependent variable, $F(6, 15) = 30.416$, and $p (0.000) < 0.05$, This indicates that the regression model is a good fit of the data.

Table 19. *Statistical significance of the independent variables.*

Coefficients[a]								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	482948082.628	0.520		.636	.0534	-1135811218.459	2101707383.714
	PV	-1.504	1.474	-.370	-1.020	.032	-4.647	1.639
	AC	2.318	.509	1.474	4.552	.000	1.233	3.403
	EV	-.923	.564	-.210	-1.636	.012	-2.125	.279
	BAC	.320	.243	.202	1.316	.020	-.198	.837
	PD	122578.252	938.713	.123	1.307	.021	-77306.342	322462.846
	VO	8668.189	109.159	.069	.809	.043	-14157.843	31494.220

a. Dependent Variable: EAC

Statistical significance of independent variables indicates that changes in the independent variables correlate with shifts in the dependent variable. Table 19 shows that the value of ($p < 0.05$) independent variables in this model is statistically significant and already appears in the prediction of the EAC model. The standard error of the coefficients in regression output is also required to be as small as possible. It reflects how wrong could be while estimating its value. In this model relative to the coefficient of the variables, their standard error is small. The prediction formula was generated by multiple regression analysis with the information provided in the coefficients under Table 19.

Statistics of regression for model EAC are generated in Table 19, which can be formulated as Equation (19) below:

$$\text{Predicted EAC} = 482948082.628-1.504PV+2.318AC-0.923EV+0.320BAC+122578.252PD+8668.189VO \quad (19)$$

Standardized coefficients are called beta weights that rank the predictor variables based on their contribution (irrespective of sign) in explaining the outcome variable. In this model, AC and BAC are the highest contributing predictor variables to explain EAC with Standardized coefficient values of 1.474 and 0.202 respectively.

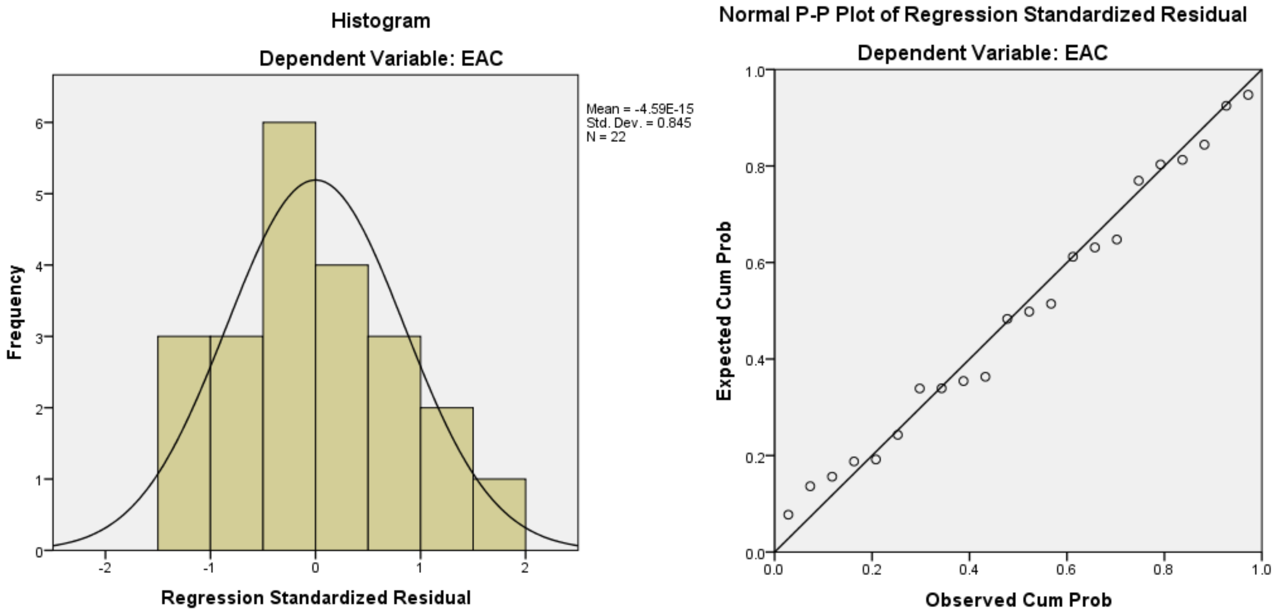


Figure 9. Histogram of Residuals for the EAC regression model and Plot of observed EAC against expected EAC for P3.

Project Four: Estimate at compilation Regression Model

The correlation coefficient (R) and coefficient of determination (R^2) for project four are .924, and .854 respectively which indicates an excellent level of prediction. The results of a correlation coefficient, and coefficient of the determination indicate a positive relationship and a high correlation between the dependent and the independent variable. Adjusted R^2 is another important factor that provides a more realistic estimate of the model accuracy than R^2. The value of adjusted R2 is .850 indicating that 85.0% variance in the dependent variable that is explained by the independent variables. The standard error of the estimate is a statistical measure used to assess the accuracy of predictions made by a regression model. The standard error of a model fit is a measure of the precision of the model. It shows how wrong one could use the regression model to make predictions or to estimate the dependent variable or variable of interest. In this estimate of the estimate at compilation regression model, the model will be wrong by 0.039 which is a small amount. The regression statistics of the estimate at compilation regression model for project four (EAC-Model) are shown in Table 20.

Table 20. Entered/Removed Variables for the EAC Regression model.

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	VO, BAC, PD, AC, EV, PV[b]		Enter

a. Dependent Variable: EAC

b. All requested variables entered.

Table 21. *Model Summary.*

Model Summary[b]									
Model	**R**	**R Square**	**Adjusted R Square**	**Std. Error of the Estimate**	**Change Statistics**				
					R Square Change	**F Change**	**df1**	**df2**	**Sig. F Change**
1	.924a	.854	.850	0.039414	.999	3116.087	6	15	.000

a. Predictors: (Constant), VO, BAC, PD, AC, EV, PV

b. Dependent Variable: EAC

Table 22. *Statistical significance of the model.*

ANOVA[a]						
Model		**Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**
1	Regression	4623050195497284.000	6	770508365916214.000	3116.087	.000[b]
	Residual	3709018493711.425	15	247267899580.762		
	Total	4626759213990995.000	21			

a. Dependent Variable: EAC

b. Predictors: (Constant), VO, BAC, PD, AC, EV, PV

Table 22 shows that the F-ratio tests whether the overall regression model is a good fit for the data and also shows that the independent variables statistically significant prediction of the dependent variable, F (6, 15) = 3116.087, and p (0.000) < 0.05, This indicates that the regression model is a good fit of the data.

Table 23. *Statistical significance of the independent variables.*

Coefficients[a]								
Model		**Unstandardized Coefficients**		**Standardized Coefficients**	**t**	**Sig.**	**95.0% Confidence Interval for B**	
		B	**Std. Error**	**Beta**			**Lower Bound**	**Upper Bound**
1	(Constant)	15043744.089	10109341.124		1.488	.015	-6503806.455	36591294.634
	PV	.047	.028	.042	1.673	.011	-.013	.107
	AC	1.105	.046	.562	23.823	.000	1.006	1.204
	EV	-1.420	.059	-.258	-23.967	.000	-1.546	-1.293
	BAC	1.143	.015	.734	74.736	.000	1.110	1.175
	PD	1141.424	845.205	.012	1.350	.019	-660.089	2942.936
	VO	11.328	54.607	.002	.207	.838	-105.064	127.721

a. Dependent Variable: EAC

Statistical significance of independent variables indicates that changes in the independent variables correlate with shifts in the dependent variable. Table 23 shows that the value of ($p < 0.05$) independent variables in this model is statistically significant and already appears in the prediction of the EAC model except for variation order. The standard error of the coefficients in regression output is also required to be as small as possible. It reflects how wrong could be while estimating its value. In this model relative to the coefficient of the variables, their standard error is small. The prediction formula was generated by multiple regression analysis with the information provided in the coefficients under Table 23.

Statistics of regression for model EAC are generated in Table 4, which can be formulated as Equation (20) below:

$$\text{Predicted EAC} = 15043744.089+0.047PV+1.105AC-1.420EV+1.143BAC+1141.424PD \tag{20}$$

Standardized coefficients are called beta weights that rank the predictor variables based on their contribution (irrespective of sign) in explaining the outcome variable. In this model, BAC and AC are the highest contributing predictor variables to explain EAC with Standardized coefficient values of 0.734 and 0.562 respectively.

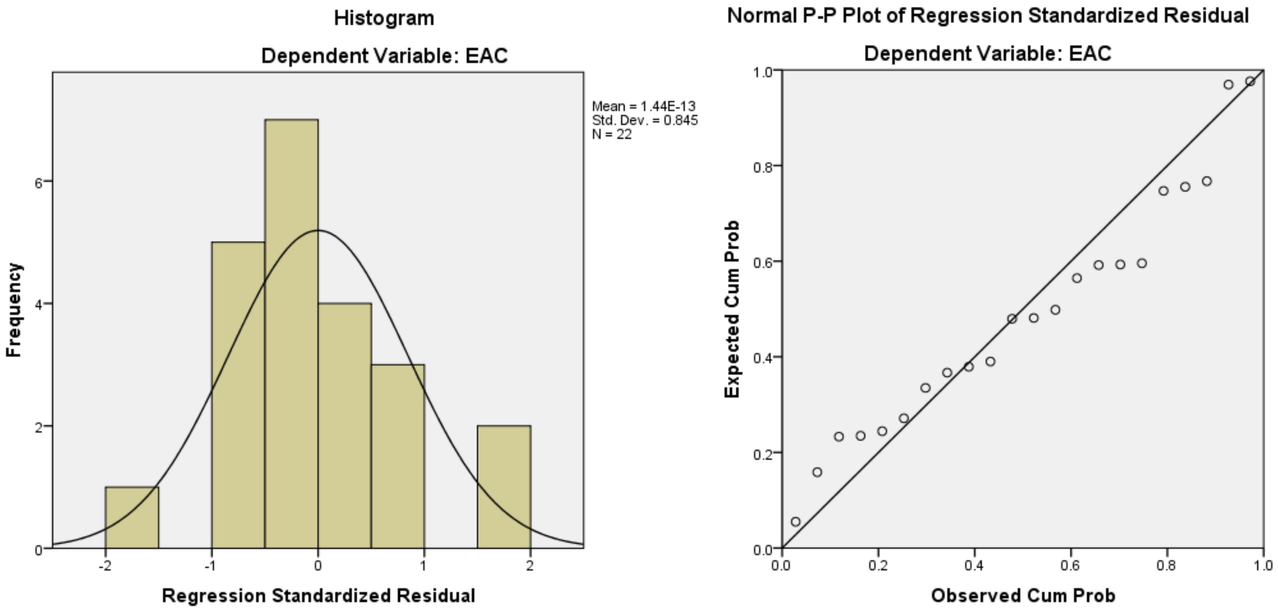


Figure 10. *Histogram of Residuals for the EAC regression model and Plot of observed EAC against expected EAC for P4.*

Project Five: Estimate at compilation Regression Model

The correlation coefficient (R) and coefficient of determination (R^2) for project five are .904, and .817 respectively which indicates an excellent level of prediction. The results of a correlation coefficient and coefficient of determination indicate a positive relationship and a high correlation between the dependent and the independent variable. Adjusted R^2 is another important factor that provides a more realistic estimate of the model accuracy than R^2. The value of adjusted R^2 is .815 indicating that 81.5% variance in the dependent variable is explained by the independent variables. The standard error of the estimate is a statistical measure used to assess the accuracy of predictions made by a regression model. The standard error of a model fit is a measure of the precision of the model. It shows how wrong one could use the regression model to make predictions or to estimate the dependent variable or variable of interest. In this estimate of the estimate at compilation regression model, the model will be wrong by 0.0475 which is a small amount. The regression statistics of the estimate at compilation regression model for project five (EAC-Model) are shown in Table 24.

Table 24. *Entered/Removed Variables for the EAC Regression model.*

Variables Entered/Removed[a]			
Model	**Variables Entered**	**Variables Removed**	**Method**
1	VO, PD, AC, EV, PV, BAC[b]	.	Enter

a. Dependent Variable: EAC

b. All requested variables entered.

Table 25. Model Summary.

Model Summary[b]									
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.904a	.817	.815	0.0475	1.000	221420.057	6	15	.000

a. Predictors: (Constant), VO, PD, AC, EV, PV, BAC

b. Dependent Variable: EAC

Table 26. Statistical significance of the model.

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	223095267563972544.000	6	37182544593995424.000	221420.057	.000[b]
	Residual	2518914400554.361	15	167927626703.624		
	Total	223097786478373088.000	21			

a. Dependent Variable: EAC

b. Predictors: (Constant), VO, PD, AC, EV, PV, BAC

Table 26 shows that the F-ratio tests whether the overall regression model is a good fit for the data and also shows that the independent variables statistically significant prediction of the dependent variable, $F (6, 15) = 221420.057$, and $p (0.000) < 0.05$, This indicates that the regression model is a good fit of the data.

Table 27. Statistical significance of the independent variables.

Coefficients[a]								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-71941579.660	1827512.822		-39.366	.000	-75836831.034	-68046328.286
	PV	-.003	.005	-.003	-.496	.027	-.014	.009
	AC	.003	.006	.003	.603	.050	-.009	.016
	EV	-1.627	.009	-1.686	-185.542	.000	-1.646	-1.608
	BAC	2.574	.010	2.664	259.149	.000	2.552	2.595
	PD	676.353	641.856	.001	1.054	.030	-691.730	2044.437
	VO	-16.430	29.237	-.001	-.562	.042	-78.746	45.886

a. Dependent Variable: EAC

Statistical significance of independent variables indicates that changes in the independent variables correlate with shifts in the dependent variable. Table 27 shows that the value of ($p < 0.05$) independent variables in this model is statistically significant and already appears in the prediction of the EAC model. The standard error of the coefficients in regression output is also required to be as small as possible. It reflects how wrong could be while estimating its value. In this model relative to the coefficient of the variables, their standard error is small. The prediction formula was generated by multiple regression analysis with the information provided in the coefficients under Table 27.

Statistics of regression for model EAC are generated in Table 4, which can be formulated as Equation (21) below:

$$\text{Predicted EAC} = -71941579.660-0.003PV-0.003AC-1.627EV+2.574BAC+676.353PD-16.430VO \tag{21}$$

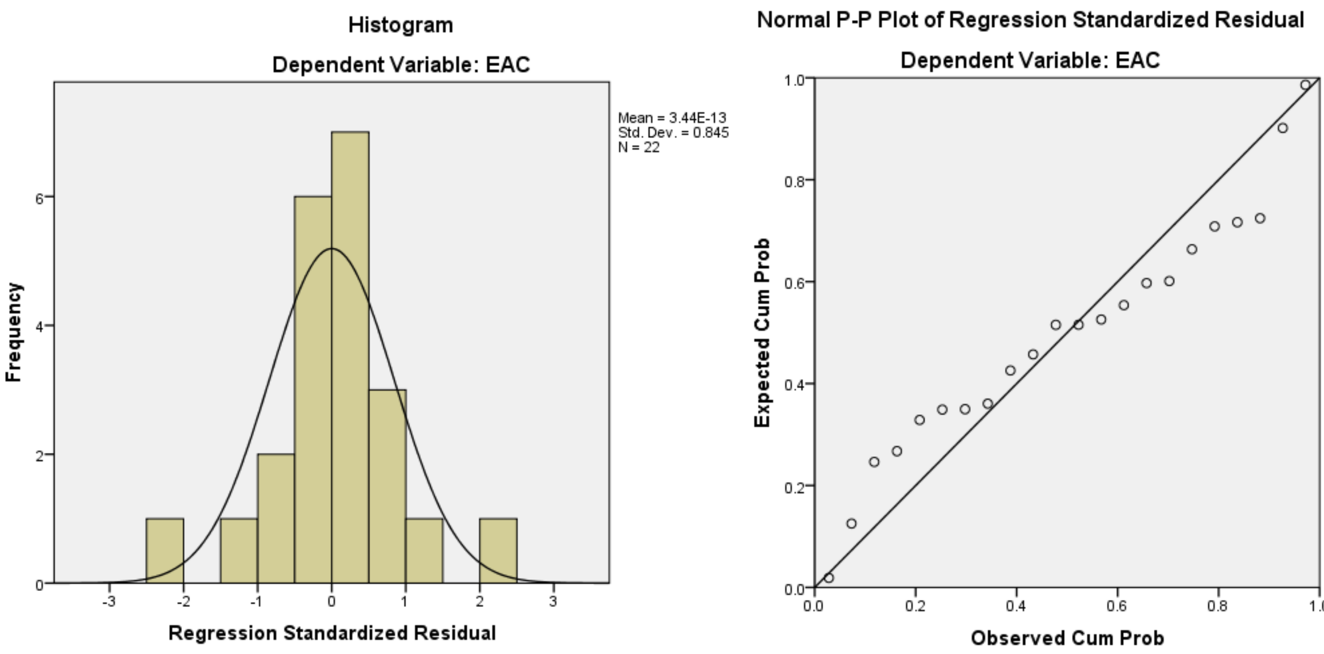


Figure 11. *Histogram of Residuals for the EAC regression model and Plot of observed EAC against expected EAC for P5.*

5. Discussion of Results

This study were carried out by a detailed case study analysis of five building construction projects to find out the most significant factors contributing to the prediction of Estimate at Compilation as output of the fuzzy controller by using MATLAB R2014a, and the Regression model. In the fuzzy logic controller, all membership functions are considered triangular membership functions with five segments. Assigning of membership function range and making the rule determine the dependability of the result produced by the rule viewer. In this study planned value, earned value, actual cost, budget at completion, project duration, and variation order were used as input variables, and estimate at compilation was used as output in terms of membership function value [0, 100] can be obtained. Similarly, the regression model affirms that it possesses a proven ability to estimate or predict the response parameters with an acceptable accuracy degree by using SPSS software. Historical data in this technique was built up on five sets of public building projects database gathering between 2021 and 2024 through the case study method as a data collection instrument from project managers, contractors, and consultants.

According to the results of numerical data analysis performed in Section 4, the summary of computing estimate at compilation by regression model for data verification and validation of estimating models has been explained as follows.

5.1. Validation and Verification of Estimate at Compilation Regression Model

In this study, validation and verification of the estimate at compilation model was checked and verified through fuzzy logic controller, and regression model. As was observed in the results of numerical data analysis performed in Section 4, planned value, earned value, actual cost, budget at completion, project duration, and variation order were used as input variables, and estimate at compilation was used as output in terms of membership function value [0, 100] can be obtained. The input values [76.97, 81.46, 81.46, 76.97, 72.47, 70.22] produce 70% of the estimate at compilation which is quite an acceptable value. Moreover, the regression model statistical approach is used to determine the model's accuracy through coefficient of determination (R2), and coefficient of correlation (R). This statistical analysis brought a prediction model of the Estimate at compilation regression model by considering PV, AC, EV, BAC, PD, and VO as an input variable.

The results of the coefficient of correlation for each project were presented under the model summary of the regression model which are 95.70% for project 1, 99.90% for project 2, 96.10% for project 3, 92.40% for project 4, and 90.40% for project 5, which is viewed as extremely aloft correlation. Similarly, the coefficient of determination shows the extent of the variety in input variable from an output variable that the model has 91.70% for project 1, 99.99% for project 2, 92.40 for project 3, 85.40% for project 4, and 81.70% for project 5. As a result, it can be concluded that these models show excellent agreement with the actual measurements.

5.2. Validation of Estimate at Compilation Multiple Regression Model

The numerical results obtained from regression analysis for model validation were performed in different tools and techniques for acceptable descriptions of the historical data. In this study tables of critical value, and the "t" test were us to validate the proposed model.

1) Using a table of critical values

The 95% critical value of the sample correlation coefficient as shown in Table 28 may be utilized to assign a valid idea of weather the calculated value of R is significant or not significant. If (R) is not between the negative and positive critical values then the correlation coefficient is significant. If (R) is significant can be used to forecasting. Value (R) computed value of estimate at compilation regression model for each building construction project is 95.70% for project 1, 99.90% for project 2, 96.10% for project 3, 92.40% for project 4, and 90.40% for project 5.

In this study, five sets of real-life building construction projects were taken for conducting the study. Thus, the Degree of freedom (df=n-2=5-2=3). The correlation between degrees of freedom and critical values is presented in Table 28.

Table 28. Critical Values: Pearson Correlation [11].

Degrees of freedom n-2	Critical Value (+ and -)	Degrees of freedom n-2	Critical Value (+ and -)
1	0.997	8	0.632
2	0.95	9	0.602
3	0.878	10	0.576
4	0.811	20	0.423
5	0.754	30	0.349
6	0.707	40	0.304
7	0.666	50	0.273

Based on the relationship between degrees of freedom and critical values presented in Table 28 the critical values correlated with degrees of freedom equal to 3 ranging between (±0.878). If the value (R) is less than the negative critical value or (R) greater than the positive critical value, then the correlation coefficient (R) is significant. Since, the correlation coefficients (R) were 95.70% for project 1, 99.90% for project 2, 96.10% for project 3, 92.40% for project 4, and 90.40% for project 5 which is greater than 0.878. Thus, the correlation coefficient values for all projects were greater than the critical value. Therefore, the correlation coefficient is significant and the multiple linear regression models can be applied for the prediction of the Estimate at Compilation regression model.

2) "t" test:

The "t" test is another statistical test used to test whether the coefficient of correlation is statistically significant or not in which the value of "t" is computed by the following Equation [11].

$$t = \frac{R * \sqrt{N-2}}{\sqrt{1-R^2}} \tag{22}$$

$$t_{EAC_1} = \frac{0.957 * \sqrt{5-2}}{\sqrt{1-0.957^2}} = 5.714$$

$$t_{EAC_2} = \frac{0.999 * \sqrt{5-2}}{\sqrt{1-0.999^2}} = 38.70$$

$$t_{EAC_3} = \frac{0.961 * \sqrt{5-2}}{\sqrt{1-0.961^2}} = 6.018$$

$$t_{EAC_4} = \frac{0.924*\sqrt{5-2}}{\sqrt{1-0.924^2}} = 4.185$$

$$t_{EAC_5} = \frac{0.904*\sqrt{5-2}}{\sqrt{1-0.904^2}} = 3.662$$

If the computed value of "t" is greater than the table value (t-distribution), the correlation is taken as significant. For 95% confidence and df = 5-1= 4, the t- table = 2.776 and t computed value for EAC value were equal to 5.714, 38.70, 6.018, 4.185 and 3.662. Therefore, t computed t-table. This indicates that the correlation coefficient is highly significant, and the multiple linear regression models can be applied for the prediction of Estimate at Compilation.

5.3. Verification of Estimate at Compilation Regression Model

The summary of computing EAC by MLR for verification of predicting models has been explained in Table 29. The first column indicates project detail, the second column shows the type of project, the third column indicates the location of the project, column four shows the actual EAC obtained from a public building project under construction in Ethiopia, and the last column presents the predicted value of EAC after applying MLR equation on them, where MLR equation has gotten through SPSS program. The comparison between the actual EAC and predicted EAC is shown in Table 29.

Table 29. *Actual EAC, and Predicted EAC for public building construction projects.*

Month	Project 1		Project 2		Project 3	
	Actual EAC	Predicted EAC	Actual EAC	Predicted EAC	Actual EAC	Predicted EAC
1	205,605,835.6	206,052,700.5	272,219,784.0	253,365,297.4	714,039,611.9	753,021,816.0
2	209,746,810.1	205,848,350.1	334,457,291.5	340,310,946.1	701,306,250.6	705,901,707.5
3	215,896,689.3	212,019,912.2	341,938,317.7	346,585,076.2	718,686,631.0	693,625,099.4
4	227,962,412.9	227,755,161.3	357,720,699.3	361,080,769.5	856,641,729.7	892,589,105.2
5	236,168,424.7	237,048,299.4	367,309,890.3	370,331,842.6	896,898,910.1	930,981,562.1
6	242,429,270.3	243,710,809.5	373,405,540.6	376,105,263.3	928,972,408.1	939,001,673.3
7	246,053,665.8	244,362,819.8	379,470,201.4	380,268,733.0	945,333,383.3	941,022,149.4
8	262,989,930.3	263,932,508.6	389,223,427.6	389,341,478.6	976,260,411.1	960,125,396.5
9	262,069,960.3	261,218,934.8	445,024,815.4	447,592,668.2	993,886,624.6	979,500,968.8
10	277,066,162.3	280,011,233.7	448,024,517.0	450,927,333.2	1,036,800,659.1	1,024,725,225.4
11	291,895,432.8	295,997,146.0	460,077,572.4	461,496,030.5	1,054,867,528.0	1,021,277,314.8
12	290,683,321.9	293,933,717.2	467,632,115.0	470,789,324.5	1,072,934,396.8	1,013,209,600.3
13	297,891,146.9	299,687,165.6	555,884,326.1	551,173,249.4	1,120,789,513.0	1,084,292,005.7
14	316,170,908.6	314,526,160.7	584,216,034.8	581,715,813.1	1,115,965,386.0	1,086,743,836.6
15	339,211,645.8	340,940,351.7	620,444,213.1	620,627,503.8	1,111,730,016.3	1,136,783,863.6
16	356,789,774.4	357,534,077.6	628,150,206.8	626,707,211.2	1,129,358,149.0	1,054,176,394.2
17	371,158,529.2	370,884,973.0	660,075,038.0	661,133,387.2	1,040,590,966.9	1,036,636,331.2
18	384,994,158.1	384,341,150.3	676,070,231.0	675,521,862.2	1,046,361,238.4	1,041,489,392.2
19	384,994,158.1	384,674,836.0	694,103,196.5	694,188,161.6	1,040,360,909.2	1,096,713,790.8
20	390,494,074.6	391,279,729.9	693,138,051.8	691,297,552.3	1,088,613,001.6	1,108,116,357.6
21	414,160,382.2	411,866,106.5	707,005,656.3	706,746,582.9	1,042,050,145.7	1,085,167,992.9
22	414,160,382.2	410,966,932.1	709,240,728.0	707,525,767.5	1,081,688,935.8	1,129,035,222.6

Table 29. Continued.

Month	Project 4		Project 5	
	Actual EAC	Predicted EAC	Actual EAC	Predicted EAC
1	160,991,139.6	159,866,837.0	615,121,684.6	614,835,410.3
2	187,659,811.0	188,517,800.2	628,494,100.1	628,291,602.9
3	183,173,506.8	183,257,749.1	652,373,669.7	652,285,466.1
4	195,983,397.3	196,323,972.7	685,739,542.8	685,943,046.1
5	211,483,648.3	212,004,686.6	689,239,873.3	689,500,103.8
6	190,951,034.7	190,862,693.0	700,489,957.6	700,816,117.7
7	191,571,391.8	191,345,542.4	726,998,530.0	727,912,700.4
8	207,523,531.2	207,094,303.0	772,026,950.9	771,067,047.7
9	205,626,090.8	205,544,832.5	737,770,618.5	737,630,275.7
10	206,770,575.1	206,649,161.3	748,394,438.5	748,465,388.7
11	209,637,667.5	209,477,443.5	757,254,882.9	757,380,666.6
12	209,637,667.5	209,423,537.8	779,970,669.7	780,231,190.5
13	201,740,969.8	201,857,173.9	805,541,249.1	804,859,941.3
14	199,511,423.6	199,662,507.1	854,992,165.3	854,826,251.3
15	199,511,423.6	199,706,700.9	884,956,771.8	884,963,075.1
16	199,511,423.6	199,795,627.3	892,343,231.3	892,363,727.6
17	199,511,423.6	199,865,627.6	896,549,409.6	896,572,473.3
18	201,858,616.8	202,230,740.0	911,014,559.6	911,132,426.6
19	199,517,039.4	199,728,133.1	911,014,559.6	911,084,015.5
20	216,953,284.1	216,940,962.0	911,014,559.6	911,073,539.3
21	224,804,395.7	224,754,725.2	911,014,559.6	911,071,687.3
22	234,734,154.6	233,752,859.9	911,014,559.6	911,024,389.7

5.4. Summary of Main Finding

The main findings of this study were comprehensive results of the fuzzy inference system and regression model. A fuzzy inference system creates a fuzzy logic model to predict estimates at compilation which might offer better accuracy. The reason why fuzzy set theory has been selected as the reliable analysis technique is that, it may handle the subjectivity that can assist project managers in estimating the future status of the project more robustly and reliably. As it was presented in the fuzzy rule viewer the input variables produce 70% of the estimate at compilation which is quite an acceptable value. Moreover, it can handle multiple inputs easily and quantify more realistically the classical problem analysis. On the other hand, prior knowledge of this theory is required. Therefore, if a case is missed, the fuzzy logic controller will not work properly. Furthermore, to produce a more robust output and validate the predicted model the regression model technique shows excellent results of prediction.

1) The most common statistical approaches employed to determine the model's accuracy are coefficient of determination (R^2), and coefficient of correlation (R). The value of the coefficient of correlation is 95.70% for Project 1, 99.90% for Project 2, 96.10% for Project 3, 92.40% for Project 4, and 90.40% for Project 5, which is

viewed as an extremely aloft correlation. Similarly, the coefficient of determination is 91.70% for Project 1, 99.99% for Project 2, 92.40 for Project 3, 85.40% for Project 4, and 81.70% for Project 5.

2) Thus, based on the results of the coefficient of correlation, and coefficient of determination the regression model maintains a more accurate and precise estimate of estimate at compilation compared to the traditional estimate at compilation, as it means reliability in the accurate predicting of the regression model.
3) In this study tables of critical value, and the “t” test were used to validate the proposed model. As it was observed in Table 28 the correlation coefficient values for all projects were greater than the tables of critical value. Therefore, the correlation coefficient is significant and the multiple linear regression models can be applied for the prediction of accurate estimates.
4) In traditional analysis of the project, data integrates only the planned value, actual cost, and earned value of the project for determining the performance indicators to evaluate their performance. Unlike the standard earned value analysis the developed model incorporates the budget at compilation, Project duration, and variation order variables in the calculation of the estimate at compilation. Thus, the demonstration of these variables in the forecasting formula reflects to increase in the accuracy and reliability of the prediction model by minimizing the standard errors associated with the estimate at compilation regression model.
5) Overall, the finding of the study revealed that the application of the regression analysis method stands for the best choice for developing the prediction model since it has more correct estimates at compilation.

6. Conclusions and Remarks

In this research cost estimate at completion prediction model was developed by using a fuzzy inference system, and regression analysis techniques through Matlab R2014a and SPSS version 23 as analysis tools.

In light of the findings and discussions the main conclusions are:

1) A fuzzy inference system creates a fuzzy logic model to predict estimates at compilation which might offer better accuracy.
2) The multiple linear regression methods can provide an important contribution to the prediction of estimates at compilation in public building construction projects.
3) In this study, Estimates at compilation models were developed using six variables as inputs which are the planned value, earned value, actual cost, budget at compilation project duration, and variation order.
4) Model validation of the proposed model was performed by using the “t” test, and the tables of critical value.
5) The computed value of the “t” test for EAC was equal to 5.714, 38.70, 6.018, 4.185, and 3.662 which is greater than the t- distribution table = 2.776. Thus, the correlation is taken as significant to predict an estimate at compilation.
6) The results of this study revealed that the value of coefficient of correlation is 95.70% for project 1, 99.90% for project 2, 96.10% for project 3, 92.40% for project 4, and 90.40% for project 5, and the coefficient of determination is 91.70% for project 1, 99.99% for project 2, 92.40 for project 3, 85.40% for project 4, and 81.70% for project 5.
7) It was found that the regression model shows excellent results in prediction.
8) It has been noticed that the regression model showed good results of estimation in terms of correlation coefficient (R) generated by multiple linear regression models for estimate at compilation for all projects.
9) At long last, a result tends to be presumed that the relationship between the dependent variable and independent variables of the built model was excellent and it shows a brilliant concurrence with the genuine prediction of the model.
10) The developed model in this study helps us to predict reliable figures in the projects which helps to make better decisions.
11) This study is an important addition to the world of knowledge, which aims to improve the appreciation of project objectives and increase confidence in expectations which can be verified statistically.
12) This developed model lies in its importance for future forecasts for other projects that have realized realistic earned value indexes that can incorporate lessons learned from similar history projects.
13) Finally, the findings of this research help academicians and practitioners with insightful information on the prediction of estimate at compilation in building projects to enhance its practicality would be the best area of investigation.

Abbreviations

AC	Actual Cost
BAC	Budget at Completion
EAC	Estimate at Completion
EVM	Earned Value Management
VAC	Variance at Completion
EV	Earned Value
SV	Schedule Variance
CV	Cost Variance
SPI	Schedule Performance Index
CPI	Cost Performance Index

Author Contributions

Girmay Getawa Ayalew: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing

Genet Melkamu Ayalew: Conceptualization, Data curation, Formal Analysis, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing

Data Availability Statement

The source of data used to support the findings of this study is available from the corresponding author upon request.

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] I. Journal *et al.*, "International journal of engineering sciences & research technology a comparison of different cost forecasting methods using earned value metrics," vol. 6, no. 6, pp. 302–319, 2017.

[2] S. M. Baqer and S. E. Rezouki, "Modeling of Estimate at Completion in Earned Value Modeling of Estimate at Completion in Earned Value," 2020, https://doi.org/10.1088/1757899X/901/1/012022

[3] S. Salikuma, M. Minu, and A. Johny, "Application of Earned Value Analysis in Analysing Project Performance," vol. 5, no. 09, pp. 459–464, 2016.

[4] T. Narbaev, "Cost Estimate at Completion Methods in Construction Projects," no. September, 2011, https://doi.org/10.13140/2.1.1898.3369

[5] T. Narbaev, "An Earned Schedule-based regression model to improve cost estimate at completion An earned schedule-based regression model to improve cost estimate at completion Timur Narbaev, Alberto De Marco International Journal Of Project Management - Elsevier - Pre," no. January 2019, 2014, https://doi.org/10.1016/j.ijproman.2013.12.005

[6] G. G. Ayalew, G. M. Ayalew, and M. G. Meharie, "Integrating exploratory factor analysis and fuzzy AHP models for assessing the factors affecting the performance of building construction projects: The case of Ethiopia Integrating exploratory factor analysis and fuzzy AHP models for assessing the factor," *Cogent Eng.*, vol. 10, no. 1, 2023, https://doi.org/10.1080/23311916.2023.2243724

[7] Shreenaath. A., Arunmozhi. S., and Sivagamasundari. R, "Prediction of Construction Cost Overrun in Tamil Nadu-A Statistical Fuzzy Approach," no. 3, pp. 267–275, 2015, [Online]. Available: www.erpublication.org

[8] N. Haloi, T. Goyal, F. Zahoor, H. Jain, and R. S. Wali, "Estimation of cost overrun in construction projects using Fuzzy Logic," vol. 23, no. 5, pp. 797–805, 2021.

[9] M. R. Feylizadeh and A. Hendalianpour, "A fuzzy neural network to estimate at completion costs of construction projects," no. April 2012, 2016, https://doi.org/10.5267/j.ijiec.2011.11.003

[10] G. Getawa Ayalew and G. M. Ayalew, "Developing fuzzy-based earned value analysis model for estimating the performance of construction projects. A case of selected public building projects in Ethiopia," *Cogent Eng.*, vol. 11, no. 1, p., 2024, https://doi.org/10.1080/23311916.2024.2348210

[11] G. M. Ayalew, M. G. Meharie, and G. G. Ayalew, "Regression Modeling for Prediction of Earned Value Indexes in Public Building Construction Projects: The case of Ethiopia Regression Modeling for Prediction of Earned Value Indexes in Public Building Construction Projects: The case of Ethiopia," *Cogent Eng.*, vol. 10, no. 1, 2023, https://doi.org/10.1080/23311916.2023.2220497

[12] D. Richard, "Analysis and application of earned value management to the Naval Construction Force," 2001.

[13] D. Marco, "ScienceDirect ScienceDirect Multiple Linear Regression Regression Model Model for for Improved Improved Project Project Cost Cost Multiple Linear Forecasting Forecasting," *Procedia Comput. Sci.*, vol. 196, no. 2021, pp. 808–815, 2022, https://doi.org/10.1016/j.procs.2021.12.079

[14] E. Computation and E. C. Studies, The Bucharest University of Economic Studies PROJECT COST ESTIMATE," vol. 52, no. 3, pp. 205–216, 2018, https://doi.org/10.24818/18423264/52.3.18.14

[15] E. G. Sinesilassie, S. Zafar, S. Tabish, and K. N. Jha, "Critical factors affecting schedule performance A case of Ethiopian public construction projects – engineers ' perspective," vol. 24, no. 5, pp. 757–773, 2025, https://doi.org/10.1108/ECAM-03-2016-0062

[16] Regression on Bridge Project 1676-1682.

[17] Shyama Salikumar and Minu Anna Johny, "Application of Earned Value Analysis in Analysing Project Performance," *Int. J. Eng. Res.*, vol. V5, no. 09, pp. 459–464, 2016, https://doi.org/10.17577/ijertv5is090341

[18] H. Khamooshi and A. Abdi, "Project Duration Forecasting Using Earned Duration Management with Exponential Smoothing Techniques," *J. Manag. Eng.*, vol. 33, no. 1, pp. 1–10, 2017, https://doi.org/10.1061/(asce)me.19435479.0000475

[19] D. Jie and J. Wei, "Estimating Construction Project Duration and Costs upon Completion Using Monte Carlo Simulations and Improved Earned Value Management," *Buildings*, vol. 12, no. 12, 2022, https://doi.org/10.3390/buildings12122173

[20] A. H. Memon, I. A. Rahman, and M. F. A. Hasan, "Significant causes and effects of variation orders in construction projects," *Res. J. Appl. Sci. Eng. Technol.*, vol. 7, no. 21, pp. 4494–4502, 2014, https://doi.org/10.19026/rjaset.7.826

[21] H. Huang and C. Ho, "Applying the Fuzzy Analytic Hierarchy Process to Consumer Decision-Making Regarding Home Stays," vol. 5, no. February 2013, 2016, https://doi.org/10.4156/ijact.vol5.issue4.119

[22] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature," *Geosci. Model Dev.*, vol. 7, no. 3, pp. 1247–1250, 2014, https://doi.org/10.5194/gmd-7-1247-2014

[23] "Using Relative Importance Index Method for Developing Risk Map in Oil and Gas Construction Projects," *J. Kejuruter.*, vol. 32, no. 3, pp. 441–453, 2020, https://doi.org/10.17576/jkukm-2020-32(3)-09

[24] G. Dreyfus, *Elementary statistics*, vol. 207. 2006. https://doi.org/10.4324/9781315697239-4.

[25] G. G. Ayalew, M. G. Meharie, and B. Worku, "A road maintenance management strategy evaluation and selection model by integrating Fuzzy AHP and Fuzzy TOPSIS methods : The case of Ethiopian Roads Authority," *Cogent Eng.*, vol. 9, no. 1, 2022, https://doi.org/10.1080/23311916.2022.2146628

[26] G. G. Ayalew, L. A. Alemneh, and G. M. Ayalew, "Exploring fuzzy AHP approaches for quality management control practices in public building construction projects Exploring fuzzy AHP approaches for quality management control practices in public building construction projects," *Cogent Eng.*, vol. 11, no. 1, p., 2024, https://doi.org/10.1080/23311916.2024.2326765

[27] T. Zsuzsanna and L. Marian, "Multiple regression analysis of performance indicators in the ceramic industry," *Procedia - Soc. Behav. Sci.*, vol. 3, no. 12, pp. 509–514, 2012, https://doi.org/10.1016/S2212-5671(12)00188-8

[28] A. Gupta, A. Sharma, and A. Goel, "Review of Regression Analysis Models," vol. 6, no. 08, pp. 58–61, 2017.

[29] R. R. R. M. Rooshdi, M. Z. A. Majid, S. R. Sahamir, and N. A. A. Ismail, "Relative importance index of sustainable design and construction activities criteria for green highway," *Chem. Eng. Trans.*, vol. 63, no. 2007, pp. 151–156, 2018, https://doi.org/10.3303/CET1863026

[30] A. Wahed, "Statistical Analysis: Internal-Consistency Reliability And Construct Validity Said Taan EL Hajjar Ahlia University," vol. 6, no. 1, pp. 27–38, 2018.

Biography

Girmay Getawa Ayalew is currently a lecturer at the Woldia Institute of Technology, Woldia University, P.O. Box 400, Woldia, Ethiopia. He is former a lecturer at the Gondar Institute of Technology, University of Gondar, P.O. Box 196, Gondar, Ethiopia. He received his BSc degree in Construction Technology and Management from Woldia University and his MSc degree in Construction Engineering and Management from the University of Gondar with first-class honours with very great distinction. He has an extensive experience in multiple publications and patents. He is participated in different peer-reviewed articles as a reviewer and was recognized as a trusted reviewer with a high level of peer review competency. Now he is 6 years of experience in teaching and researching. His research interest includes BIM, Machine Learning Algorithms, Fuzzy Logic Regression Modelling, Sustainability in Construction, Facility Management, and Project Management, Material Management and TQM.

Genet Melkamu Ayalew is currently working as a lecturer at the Woldia Institute of Technology, Woldia University, Woldia, Ethiopia. She is former a lecturer at the Gondar Institute of Technology, University of Gondar, Gondar, Ethiopia. She received her BSc degree in Construction Technology and Management from Debre Berhan University and her MSc degree in Construction Engineering and Management from the University of Gondar. Currently she has 6 years of teaching experience and actively engaged in research. She published more than five research papers in highly indexed peer-reviewed journals. Her research interest includes Project Performance, Project Management, BIM, MLA, Maintenance Management, Fuzzy AHP, and Regression Modeling.