

**ADDIS ABABA UNIVERSITY**  
**ADDIS ABABA INSTITUTE OF TECHNOLOGY**  
**SCHOOL OF CIVIL AND ENVIRONMENTAL**  
**ENGINEERING**



**Modeling Structural Works Construction Duration: A Case Study**  
**of Selected Apartment Projects in Addis Ababa**

---

A Thesis Submitted to School of Graduate Studies in Partial Fulfillment of the  
Requirements for the Degree of Master of Science in Civil Engineering  
(Construction Technology and Management)

Author: Biruk Legesse  
Advisor: Dr. Abraham Assefa

October 2020  
Addis Ababa

The undersigned have examined the thesis entitled ‘**Modeling Structural Works Construction Duration: A Case Study of Selected Apartment Projects in Addis Ababa**’ presented by **Biruk Leggesse**, a candidate for the degree of **Master of Science** and hereby certify that it is worthy of acceptance.

Dr. Abraham Assefa	_____	_____
Advisor	Signature	Date
Dr. Dereje Hailu	_____	_____
Internal Examiner	Signature	Date
Dr. Asregedew Taye	_____	_____
External Examiner	Signature	Date
_____	_____	_____
Chair person	Signature	Date

## **UNDERTAKING**

I certify that research work titled “Modeling Structural Works Construction Duration: A Case Study of Selected Apartment Projects in Addis Ababa” is my own work. The work has not been presented elsewhere for assessment. Where material has been used from other sources it has been properly acknowledged / referred.

Author: Biruk Leggesse

Signature:

## ACKNOWLEDGMENTS

First and foremost, praises and thanks to God, the Almighty, for His showers of blessings throughout my research work to complete the research successfully.

I would like to express my deep and sincere gratitude to my research advisor, Dr. Abraham Assefa, for giving me invaluable guidance throughout this research. His dynamism, vision, sincerity and motivation have deeply inspired me in many ways.

I am extremely grateful to my parents for their love, prayers, caring and sacrifices for educating and preparing me for my future. Specially, my mother Wolansa, I have no words to give praises for your everyday help. I am very much thankful for my brothers Yared and Sentayewe for supporting me in every way possible throughout my life. A special warm thanks for Mr. Alexander Menk, for finding and sending me paid journal articles which made the literature review complete. I am also, grateful to acknowledge my sisters; Eshete, Tigist, and Helen, for their encouragement throughout the years.

I would like to thank my friends specially Gelu, for his invaluable help during data collection.

My sincere thanks also goes to my beloved half Weyni, who was with me encouraging, giving me support and love all the way to the end.

Last but not least, I would like to thank Ethiopian Road Authority, for providing me this scholarship. Also, acknowledgements goes to all real estate developers who participated in the study.

Biruk Leggesse

## ABSTRACT

There is a recognized need for accurate estimation of construction duration of buildings. Uncertainties in duration estimation hindered a problem to the construction sector. With this regard, approximate estimation of the various phases of building construction indicated to minimize the uncertainty in duration estimation. The structural framework construction phase covers nearly half of the building project duration. Further, also this phase almost always present in the critical path of project schedule. With this regard, accurate and early prediction of structural frameworks duration can help mobilize and manage resources effectively. Thus, this study provides parametric models that ease this need.

Case study samples of 41 apartment residential buildings constructed and completed between a period from 2010 and 2019 were collected using structured questionnaire. 33 of the sample used to develop duration prediction models using linear regression analysis, while the remaining 8 samples used to validate developed models.

In order to develop mathematical models first, the interrelationship that exist between structural works duration influencing factors; project scope factors and quantity of work items investigated. Project scope factors; total gross floor area, and typical floor area showed a statistically significant moderate strength positive linear relation. Similar linear relations were also observed with quantity of work items; quantity of excavation, formwork, reinforcing bar, and quantity of concrete.

Simple and multiple linear regression analysis carried out resulted in models with low coefficient of determination. But, best-fit models were developed, by splitting the model building data set into small-size and large-size projects based on total gross floor area and removing outliers. More than 85 percent of significant variance in structural frame work duration by selected project scope and quantity of work item predictor variables was accounted by best-fit models.

The present study put efforts towards enhancing our understanding of relationships existing between selected predictor variables and structural works construction duration.

**Key words:** structural framework; construction duration; parametric model; quantity of work item; project scope factors

## TABLE OF CONTENTS

<b>TABLE OF CONTENTS</b>	
<b>UNDERTAKING</b> .....	<b>II</b>
<b>ACKNOWLEDGMENTS</b> .....	<b>III</b>
<b>ABSTRACT</b> .....	<b>IV</b>
<b>TABLE OF CONTENTS</b> .....	<b>V</b>
<b>LIST OF TABLES</b> .....	<b>VIII</b>
<b>LIST OF FIGURES</b> .....	<b>X</b>
<b>LIST OF ABBREVIATIONS</b> .....	<b>XI</b>
<b>CHAPTER 1 INTRODUCTION</b> .....	<b>1</b>
1.1 Background of the Study .....	1
1.2 Problem Statement .....	2
1.3 Research Significance .....	3
1.4 Research Objectives .....	4
1.4.1 General objectives.....	4
1.4.2 Specific objectives .....	4
1.5 Research Questions .....	4
1.6 Research Hypotheses .....	5
1.7 Research Limitations.....	5
1.8 Thesis Organization .....	6
<b>CHAPTER 2 LITERATURE REVIEW</b> .....	<b>7</b>
2.1 Introduction .....	7
2.2 The Big Picture: Project Schedule Management .....	8
2.2.1 Basic definitions and introductions.....	9
2.2.2 PMBOK approach to schedule management.....	11
2.3 Activity Duration Estimates .....	13
2.3.1 Activity duration defined.....	13
2.3.2 PMBOK approach to duration estimation.....	13
2.4 Building Construction Duration Influencing Factors.....	15

2.5	Modeling Construction Duration of Building Projects .....	20
2.5.1	Linear Regression Approach for Modeling Construction Duration .....	20
2.5.2	Other approaches for modeling construction duration.....	35
2.6	Modeling Construction Duration in Ethiopia .....	41
2.7	Modeling Structural Works Construction Duration.....	43
2.8	Construction Duration Model Prediction Performance Measures .....	47
2.9	Summary of Literature Review and Research Gap Identified.....	50
2.9.1	Summary of literature review .....	50
2.9.2	Research gap identified .....	52
<b>CHAPTER 3</b>	<b>RESEARCH METHODOLOGY .....</b>	<b>53</b>
3.1	Introduction .....	53
3.2	Research Design.....	53
3.2.1	Type of research.....	53
3.3	Population of the Study .....	54
3.4	The Study Sample and Sample Size Determination .....	54
3.4.1	Study sample.....	54
3.4.2	Sample size determination.....	55
3.5	Identified Variables .....	56
3.5.1	Operational definitions of identified variables .....	59
3.6	Data Collection .....	62
3.6.1	Instrument.....	62
3.6.2	Respondents size determination .....	63
3.6.3	Sampling Procedure of Respondents .....	63
3.7	Statistical Data Analysis Approach.....	65
3.7.1	Data preparation.....	65
3.7.2	Statistical methods used .....	65
3.7.3	Hypotheses testing .....	66
3.7.4	Validity of results.....	67
3.8	Survey Distribution and Response Statistics .....	69
<b>CHAPTER 4</b>	<b>RESULTS AND DISCUSSIONS .....</b>	<b>71</b>

4.1	Introduction to Results and Discussions Section .....	71
4.2	General Description of Data Sets.....	72
4.2.1	Data set.....	72
4.2.2	Model building data set.....	72
4.2.3	Model validation data set .....	73
4.3	Correlation Analysis, Results and Discussions .....	74
4.3.1	Structural works duration correlation with project scope factors.....	74
4.3.2	Structural works duration correlation with quantity of work items.....	79
4.4	Analysis, Results and Discussions of Hypotheses .....	82
4.4.1	Hypothesis of linear association with project scope factors.....	82
4.4.2	Hypothesis of linear association with quantity of work items.....	86
4.5	Structural Framework Construction Duration Prediction Models .....	89
4.5.1	Simple Linear Regression (SLR) Models .....	89
4.5.2	Multiple Linear Regression (MLR) Models.....	93
4.5.3	Best-fit Linear Regression Models .....	97
4.6	Comparison of Different Construction Methods .....	101
4.6.1	Non-parametric tests for comparing difference between groups.....	101
<b>CHAPTER 5 CONCLUSIONS AND RECOMMENDATIONS.....</b>		<b>106</b>
5.1	Conclusions .....	106
5.2	Recommendations.....	108
5.2.1	Recommendations for project stakeholders .....	108
5.2.2	Future research areas.....	108
<b>REFERENCES.....</b>		<b>109</b>
<b>APPENDICES .....</b>		<b>119</b>

## LIST OF TABLES

Table 2-1 Estimate Activity Durations Process adopted from (Sanghera, 2019) and (PMI, 2017).....	14
Table 2-2 Project related factors that affect duration of construction building projects ..	18
Table 2-3 AIQS survey values of K adopted from (Bromilow, et al., 1980 ) .....	25
Table 2-4 Time-cost performance model for selected projects in Hong Kong (Chan & Kumaraswamy, 1995) .....	26
Table 2-5 Selected examples of BTC model validations for building projects .....	27
Table 2-6 ANN terminologies and their statistical counterparts adopted from (Maier & Dandy, 2000) as cited in (Guerrero, 2016).....	38
Table 2-7 Comparison of previous structural work duration estimation models .....	46
Table 3-1 Identified dependent variables of the study .....	58
Table 3-2 Identified independent variables of the study .....	58
Table 3-3 structural framework duration identified influencing factors categorization ..	59
Table 3-4 Qualitative variables used with their identified measurement categories .....	61
Table 3-5 Calculation of size of respondents (real estate developers).....	63
Table 3-6 Regression diagnostics used for models, slightly modified from (Guerrero, 2016).....	68
Table 3-7 Summary of questionnaire survey response rate.....	70
Table 4-1 Descriptive summary of dependent variables of model building data set .....	72
Table 4-2 Descriptive summary of project scope variables of model building data set ..	72
Table 4-3 Descriptive summary of quantity of works variables of model building data set .....	73
Table 4-4 Descriptive summary of validation data set.....	73
Table 4-5 Pearson’s correlations between dependent variables & project scope factors.	76
Table 4-6 Pearson’s correlation of transformed $D_{Total}$ with project scope factors .....	76
Table 4-7 Pearson’s correlation of transformed $D_{Super}$ with project scope factors .....	77
Table 4-8 Pearson’s correlation of transformed $D_{Sub}$ with project scope factors .....	77
Table 4-9 Suggestion on strength of linear correlation adopted from (Jackson, 2009) ...	78
Table 4-10 Pearson’s correlations between dependent variables & quantity of work items .....	80
Table 4-11 Pearson’s correlation of transformed $D_{Total}$ with quantity of work items ..	80
Table 4-12 Pearson’s correlation of transformed $D_{Super}$ with quantity of work items	81

Table 4-13 Pearson’s correlation of transformed <i>D_Sub</i> with quantity of work items ...	81
Table 4-14 Statistical results of <i>D_Total</i> for Hypothesis 1 (H1).....	83
Table 4-15 Statistical results of <i>D_Super</i> for Hypothesis 1 (H1).....	84
Table 4-16 Statistical results of <i>D_Sub</i> for Hypothesis 1 (H1).....	85
Table 4-17 Statistical results of <i>D_Total</i> for Hypothesis 2 (H2).....	87
Table 4-18 Statistical results of <i>D_Super</i> for Hypothesis 2 (H2).....	87
Table 4-19 Statistical results of <i>D_Sub</i> for Hypothesis 2 (H2).....	88
Table 4-20 SLR models between dependent variables and project scope factors.....	90
Table 4-21 SLR models between dependent variables and quantity of work items.....	90
Table 4-22 Comparison of SLR models for <i>D_Total</i> .....	92
Table 4-23 Comparison of SLR models for <i>D_Super</i> .....	93
Table 4-24 Comparison of SLR models for <i>D_Sub</i> .....	93
Table 4-25 Predictive Performance of Multiple Linear Regression (MLR) Models.....	94
Table 4-26 Statistical data of the regression coefficients of MLR models.....	95
Table 4-27 Predictive models comparison and validation statistics.....	96
Table 4-28 Predicted durations of validation data set projects.....	97
Table 4-29 Data splitting based on gross floor area and outliers removed.....	98
Table 4-30 Best-fit models statistical performance for small-size apartments.....	99
Table 4-31 Best-fit models statistical performance for large-size apartments.....	100
Table 4-32 Validation of best-fit models.....	100
Table 4-33 Frequency table of construction method factor groups.....	101
Table 4-34 Mann-Whitney U Test Statistics for Foundation type.....	103
Table 4-35 Mean Rank values of durations in slab type categories.....	104
Table 4-36Kruskal-Wallis test statistic for slab types.....	105

## LIST OF FIGURES

Figure 2-1 General Scope of the study.....	8
Figure 2-2 Schedule Management Processes (Adopted from Sanghera, 2019) .....	12
Figure 2-3 Operating diagram of an Artificial Neuron adopted from (Guerrero, 2016)..	37
Figure 3-1 Study sample composition and sample size .....	56
Figure 3-2 Classification of variables used for the study .....	57
Figure 3-3 Operational classification of structural work groups. ....	60
Figure 4-1 Matrix scatter plot together with histogram <i>D_Total</i> variable .....	75
Figure 4-2 Matrix scatter plot together with histogram <i>D_Super</i> variable .....	79

## LIST OF ABBREVIATIONS

AI	Artificial Intelligence
ANN	Artificial Neural Network
AIQS	Australian Institute of Quantity Surveyors
BaTCoDA	Building and Transport Construction Design Authority
BTC	Bromilow's Time-Cost model
CSIRO	Commonwealth Scientific and Industrial Research Organization
ERA	Ethiopia Road Authority
GFA	Gross floor area
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MEP	Mechanical, Electrical, and Plumbing
MLR	Multiple linear regression
MSE	Mean square error
PRINCE2	PRoject IN Controlled Environment
PMBOK	Project Management Body of Knowledge
PMI	Project Management Institute
RMSE	Root-mean-square error
SLR	Simple linear regression
TCR	Time –cost relationship

## CHAPTER 1

## INTRODUCTION

### 1.1 Background of the Study

Timely completion of a project is considered by many as a measure of success. Projects completed within planned time give satisfaction to project stakeholders. However, delays are very common in construction industry. Poor performance by the contractor is one of the main causes sited to cause delay. A key view presented by Mackova, et al., (2017) indicated that, inaccurate estimation of construction durations as one of the main causes for project delay (Mackova, et al., 2017). An accurate prediction of construction duration and budget is valuable in planning and controlling project. It can be implied that tools assisting for estimation of construction time at earlier stages of construction planning are of a vital help for all involved in the project (client, contractor, consultant, etc.). In earlier project initiation phase where, unavailability of detailed project data casts a problem in deriving a more accurate duration estimates.

It was back in 1960s where attempts to predict construction duration and cost using project specific data began (Khosrowshahi & Kaka, 1996). The development of mathematical models, that assist in prediction of duration have been interest for many researchers and practitioners. For building projects specifically mathematical models that address different duration influencing factors developed. Project scope factors, that describe the physical size of building are used for this purpose. A recent study discovered that, the quantity of work item in a given project clearly determines the required duration to complete it but previous studies don't incorporate this significant factor in predictive models (Mensah, et al., 2017).

Estimating the construction durations of the various phases of construction process during the early design stage using basic project scope data will assist both the client and contractor in having an achievable plan. Chan and Kumaraswamy (1996b) indicated that for a better analysis and drive a powerful predictive model of construction duration, the construction process can be divided into principal phases (Chan & Kumaraswamy, 1999b). The construction stage of a building project can be broadly divided into four main phases: site preparation, excavation, structural work (underground/aboveground), and finishing work. All the four phases are critical and must be planned and coordinated as precisely as

possible: thus, the total time required to complete each phase is an important factor to consider during the planning phases of construction (Lee, et al., 2009).

## **1.2 Problem Statement**

Urbanization together with increased population in capital cities lead to the adoption of high rise buildings for housing purpose. In high rise apartment projects, the structural work packages are the most critical and always equipment, labour and material intensive part of a project. It is also clear for project stakeholders that a building with a sound and firm structure is the base for the whole functionality of a building.

The importance of the structural framework in a building construction is understandable. The importance came from the very fact that all the building load for which the building is designed to resist for is carried by it. Both the design and construction of the structural framework require care any negligence or mistake in worst case scenario might lead to the collapse of the entire building. The construction of the structural framework alone occupies approximately 50 percent of the entire construction duration of the building (Bang, et al., 2011). The increase in population together with scarce land resource made countries to shift to high-rise building projects than the luxury villa housing typologies. The Upper structure work of this high-rise buildings is an element of activity that has the greatest weight with a percentage value of 21percent of the total work weight (Husin, 2019).

Despite the high importance of structural framework construction phase of a building project there exist lack of a separate treatment of this phase from researchers specifically in modeling the duration aspect.

In practice, there are two common methods of estimating project completion time: (1) according to the client's time constraints eg. Occupancy need, or (2) through a detailed analysis of work to be done and resources available, using estimates of the time requirements for each specific activity (Ng, et al., 2001). Further also the contractor during tender submission provides his work schedule that is bounded within the time frame of the client. Both contractors and consultants use their experience on previous projects for prediction of future project durations of new projects. But projects are unique in every other way from any previous one so an experience-based estimations are less accurate and might lead to time overruns.

Accurate duration prediction especially during earlier planning phase of projects is a difficult task. Mathematical duration models that use previous project statistics came to solve this problem. Several researchers attempted to discover the association between building project duration and project scope parameters as measured either in terms of total project cost or physical size. The use of cost in duration prediction models comes with time value of money problem. In Ethiopia, where there is no available data for cost index hinders the problem of comparing projects constructed in different years. This research underlines the need for a tool that would be used by both clients and contractors for forecasting construction duration of structural works in residential buildings during planning stage. During planning stage, where actual cost and duration being unlikely to be predicted accurately, such mathematical tools are helpful to have a confidence in scheduling. Models incorporating project scope and quantity of work features which are less affected by variations are stronger and tend to be more precise measure of construction time.

It is clear that the total construction duration is the cumulative sum of every activity's duration in the critical path of a schedule, structural framework construction activities almost always present in the critical path. Apart from critical path activities that determine construction duration, the factors that lead to the realization of final duration might had been highly affected by a delay in certain separate phase of the construction. For example, a delay in the finishing phase due to lack of foreign currency might push the actual duration further from what was anticipated to be, such a scenario is better understood by separating the construction duration into separate segments.

### **1.3 Research Significance**

The research findings add significant value to the existing body of knowledge. Previous building duration estimation models developed used project scope parameters which are identified as the measure of the size of a project. Cost is the major and widely used predictor variable for duration. Bromilow, (1969) identified the existence of relationship between project cost and time required to complete a given project. Current review of literature on previous studies of duration prediction models by Mensah, et al., (2017) indicated the gap in use of quantity of work items of a given project in duration prediction models. The current study developed on this identified research gap and make use of

quantity of work items in duration prediction models. The study developed models that are practical and easy for use in the construction industry with similar settings as project samples used in the study.

## **1.4 Research Objectives**

### **1.4.1 General objectives**

The main objective of this research is to develop a mathematical model for predicting the overall construction period of structural framework by taking private residential real estate apartment projects as case studies. For achieving this goal specific objectives were set.

### **1.4.2 Specific objectives**

- Identify influencing variables for predicting construction duration of overall structural work elements.
- Analyze the kind and degree of relationship between duration of structural frameworks and selected project scope parameters.
- Identify the kind and degree of relationship between duration of overall structural work elements and quantities of work items (i.e. volume of concrete, quantity of formwork and rebar)
- Generate models for predicting construction duration of substructure and super structure works.
- Identify possible differences in structural framework duration for groups of different construction method.

## **1.5 Research Questions**

To meet the general objectives of the study together with the building specific, three research questions were outlined.

RQ1. What is the degree of relationship among the dependent structural framework construction duration variables (i.e. total, substructure & super structure construction durations) and each of independent selected project scope parameters (i.e. Gross and typical floor area, total number of floors above and below ground, total and typical building height.)?

RQ2. What is the degree of association among the dependent structural framework construction duration variables (i.e. total, substructure & super structure construction durations) and each of independent selected quantities of work items (i.e. quantity of concrete, excavation, formwork and rebar)?

RQ3. Is there a difference in structural framework construction durations by selected construction method?

## **1.6 Research Hypotheses**

Using the research objective, problem statement, and research question and to arrive at a significant model for predicting construction duration of structural works with parameters that are available at earlier stage of planning, the following two research hypotheses were formulated:

H1: there is a significant relation between structural works duration and selected project scope factors (i.e. Gross and typical floor area, total number of floors above and below ground, total and typical building height.)

H2: there is a significant relation between structural works duration and selected quantity of work elements (i.e. volume of concrete, excavation, quantity of formwork and rebar).

The null hypothesis is used for conducting statistical tests at a significance alpha level of 0.05 for both hypotheses.

## **1.7 Research Limitations**

The major limitation of the study is sample size. The Corona virus pandemic hindered a serious problem on the data collection process, where offices were closed and officials were not collaborative. But samples of 41 apartment projects can be acknowledged to be sufficient for the statistical tests carried out. This is justifiable by the recommendations given by researchers on minimum sample size for conducting parametric tests (e.g. Mendenhall, et al., 1993).

The research used only a handful of quantitative predictor variables which resulted in statistically significant prediction models with low explanatory power of the dependent

variable. Resource and time limitations hindered the incorporation of qualitative predictor factors.

## **1.8 Thesis Organization**

The introduction chapter presented the background of the study, problem statement, the research objectives, research questions addressed and the hypotheses. The reminder of the thesis consists of four chapters: chapter two present the literature review, chapter three outlines the research methodology used, the result and discussion presented in chapter four and the fifth chapter provides the general conclusions and recommendations. Further, essential tables are presented in the appendices section.

## CHAPTER 2 LITERATURE REVIEW

### 2.1 Introduction

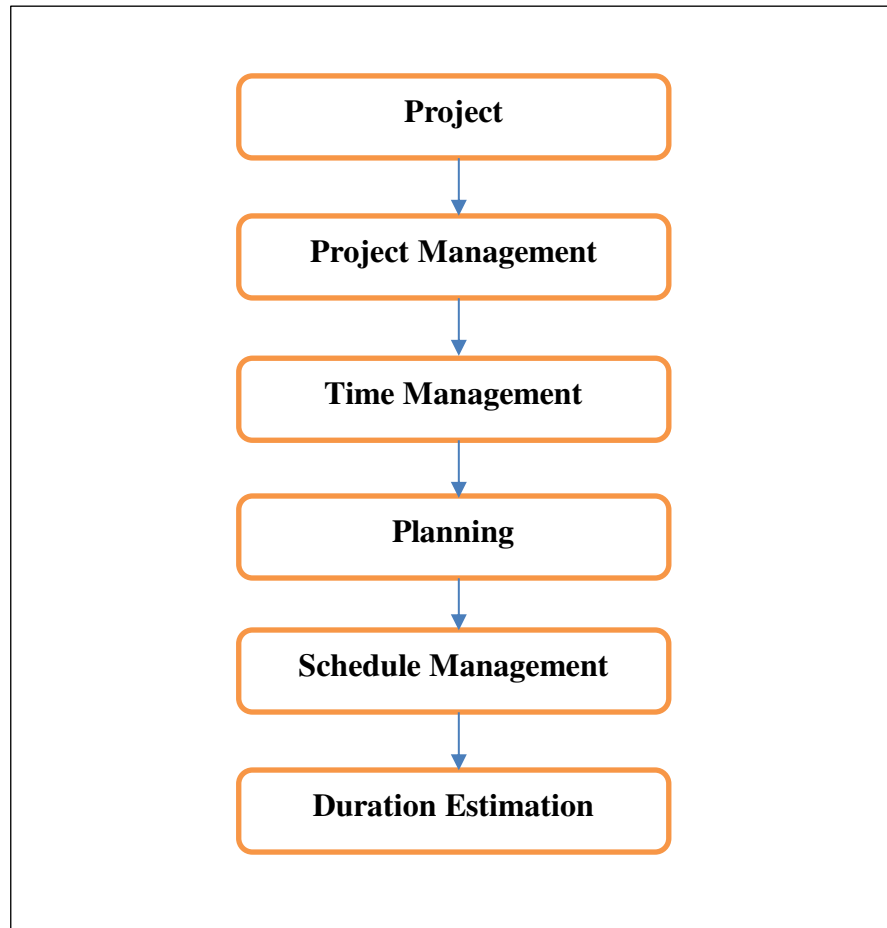
Before commencement of a research work searching for existing literature is a mandatory task. The following key words and phrases were used in searching for previous related works: duration, time, model, prediction, construction duration, construction time, forecasting building construction duration, predicting construction time, modeling construction duration, and construction time influencing factors. While the following databases were accessed using the key words and research phrases: Addis Ababa university digital library, Science direct, Taylor and Francis, Academia, ASCE Library, Semantic Scholars, Research gate; Emerald Insight, google search engine, and google scholars. Reading through the title of papers those assumed relevant were kept in a separate folder for screening. Reading through their abstract, papers were further separated into duration modeling and influencing factor identifiers and kept into two separate folders. Analysis of literatures of same category were then conducted only on papers that focus on building projects. Further, additional papers were searched using the reference list of other papers too.

The literature review section begins by showing the wider scope of the study and the limiting point of this chapter. Headings, subheadings, and sub subheadings were used to analyze the following themes:

- Building construction duration influencing factors;
- Modeling construction duration of building projects;
- Modeling construction duration in Ethiopia;
- Modeling structural works construction duration;
- Construction duration model prediction performance measures;

## 2.2 The Big Picture: Project Schedule Management

In conducting literature review it is advisable to understand where the specific concept that is going to be reviewed fall in the wider context of the study area. With this regard, a graphical diagram is developed below figure 2.1 to assist the review from a much wider scope and try to concentrate with the specific area of discussion.



**Figure 2-1 General Scope of the study**

From figure 2.1 it is clear to look where the specific topic (duration estimation) of the research falls in the project management knowledge area. The schedule management topic is found as the most suitable scope for the consideration of the study topic from a wider perspective in a much-detailed way.

The Project Management Body of Knowledge (PMBOK) is used to discuss the schedule management topic. The general scope areas illustrated in figure 2.1 are discussed in a glance first.

## 2.2.1 Basic definitions and introductions

### 2.2.1.1 Project definition

The word ‘Project’ entails different definitions and mining to different people. One of the most widely practiced management methods, Project Management Body of Knowledge (PMBOK) guide 6<sup>th</sup> edition defines project as; “*temporary endeavor undertaken to create a unique product, service, or result.*” (PMI, 2017).

The other competent management method, Project IN Controlled Environment (PRINCE2) defines ‘Project’ as “*A temporary organization that is created for the purpose of delivering one or more business products according to an agreed business case.*”

From the two widely accepted management methods definitions the following common points can be discussed;

- **Temporary**; projects are created for a temporary time period (specific start and end time), within such a period a project passes through several stages and come to an end.
- **Unique**; every project is different from its previous one in some way like, its location, team organization, stakeholders,
- **Product**; this concept identifies the necessity of a deliverable from a project. A project without a deliverable cannot be taken as a project. The deliverable can be physical product, or a service which will aid to deliver another product.

### 2.2.1.2 Project Management definition

PMI defined ‘Project Management’ as ; “*...the application of knowledge, skills, tools, and techniques to project activities to meet the project requirements.*” (PMI, 2017).

While PRINCE2 defined it as: “*...the planning, delegating, monitoring and control of all aspects of the project, and the motivation of those involved, to achieve the project objectives within the expected performance targets for time, cost, quality, scope, benefits and risk.*”

From the two definitions it is understandable that project management involves various skills, tools & techniques that help to achieve the project objectives which are defined as a unique product above in ‘Project’ definition.

Effective implementation of project management provides several benefits among which; satisfy stakeholder expectations, increase chances of success, resolve problems and issues, respond to risks in a timely manner. While in the contrary the ineffective application of project management result in; missed deadlines, cost overruns, poor quality, Unsatisfied stakeholders (PMI, 2017).

### **2.2.1.3 Time Management definition**

According to PMBOK 5<sup>th</sup> edition “*Project Time Management includes the processes required to manage the timely completion of the project*”. The recent edition of PMI’s PMBOK guide changed this knowledge area to schedule management explaining the fact that time management is much wider and project managers don’t practice time management rather engaged in schedule management.

### **2.2.1.4 Planning definition**

The terms planning and scheduling are often used synonymously but the fact is the two are separate activities (Mubarak, 2010; Baldwin & Bordoli, 2014). Baldwin & Bordoli indicated that planning is a wider context not limited to the traditional thinking of identification of time schedules. Further apart from the time aspect the concept of cost, quality, health and safety and other aspects such as design and production must be considered in planning (Baldwin & Bordoli, 2014).

PMI defined the planning process group as “*those processes required to establish the scope of the project, refine the objectives, and define the course of action required to attain the objectives that the project was undertaken to achieve.*” (PMI, 2017)

PRINCE 2 defined plan as “*a detailed proposal for doing or achieving something which specifies the what, when, how and by whom it will be achieved.*” (PRINCE2, 2017)

It is clear from the above definitions that planning comprises of a wider aspect than scheduling and the two are different terms.

### **2.2.1.5 Scheduling definition**

Mubarak, (2010) defined scheduling as the determination of the timing and sequence of operations in the project and their assembly to give the overall completion time.

Baldwin & Bordoli, (2014) listed three core questions that are answered from scheduling outcome;

- When will the work be carried out?
- How long will it take?
- What level of resources will be required?

Scheduling provides several uses among those Mubarak, (2010) identified the following ones; identification of project completion time, identification of specific activity start and end time, improve work efficiency, as a project performance control tool, and prove delay claims.

Project scheduling had been indicated by authors and practitioners as a difficult task to conduct (e.g. (Khodakaram, et al., 2007)) due to the large uncertainties involved in a project resource, duration estimate of individual tasks, and productivity rate estimates.

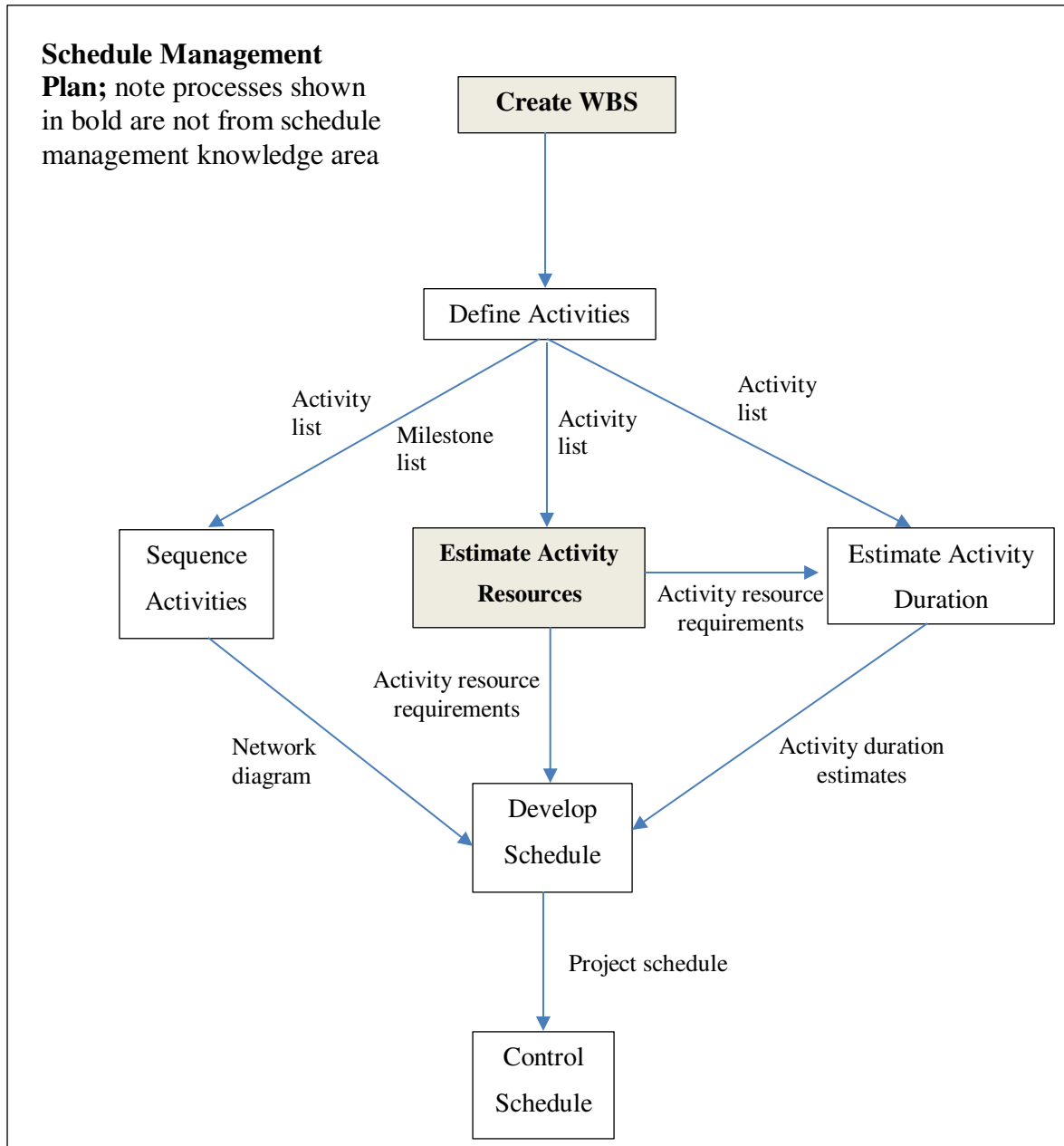
### **2.2.2 PMBOK approach to schedule management**

PMBOK guide 6<sup>th</sup> edition adopts project schedule management knowledge area replacing the previous 5<sup>th</sup> edition Time management knowledge area, to reflect that the project schedule is defined and managed during the project, whereas time is not managed (PMI, 2017).

The schedule management knowledge area is concerned with the major process groups required to manage the timely completion of the project (PMI, 2017). This knowledge area consists of six process groups;

- 1) Plan Schedule Management
- 2) Define Activities
- 3) Sequence Activities
- 4) Estimate Activity Durations
- 5) Develop Schedule
- 6) Control Schedule

Figure 2-2 below shows the schedule management processes group that lead to schedule development.



**Figure 2-2 Schedule Management Processes (Adopted from Sanghera, 2019)**

The defining tasks of the schedule management process groups are listed herein below;

- 1) **Plan Schedule Management**—The process of establishing the policies, procedures, and documentation for planning, developing, managing, executing, and controlling the project schedule.
- 2) **Define Activities**; identifying and documenting the specific actions to be performed to produce the project deliverables.

- 3) **Sequence Activities**; identifying and documenting relationships among the project activities.
- 4) **Estimate Activity Durations**; estimating the number of work periods needed to complete individual activities with the estimated resources.
- 5) **Develop Schedule**; developing the project schedule by analyzing activity sequences, durations, resource requirements, and schedule constraints.
- 6) **Control Schedule**; monitoring the status of the project to update the project schedule and manage changes to the schedule baseline. (PMI, 2017).

## **2.3 Activity Duration Estimates**

### **2.3.1 Activity duration defined**

PMBOK 6<sup>th</sup> edition defined Activity as *“A distinct, scheduled portion of work performed during the course of a project.”* And Activity duration as *“The time in calendar units between the start and finish of a schedule activity.”*

Duration. *“The total number of work periods required to complete an activity or work breakdown structure component, expressed in hours, days, or weeks.”*

Activity Duration Estimates. *“The quantitative assessments of the likely number of time periods that are required to complete an activity.”*

### **2.3.2 PMBOK approach to duration estimation**

As discussed in section 2.2.2 above PMBOK lists six process groups under project schedule management knowledge area, among which activity duration estimation is the one which will be reviewed here.

Activity duration is the time between the start and finish of an activity. Activity as defined earlier is the smallest manageable task in a work break down structure(WBS). Activity duration is estimated in work periods by using the Estimate Activity Durations process for the given resources assigned to the activity. A work period is a measurement of time when the work is in progress; it is measured in hours, days, or months, depending upon the size of the activity (Sanghera, 2019).

The inputs, tools & techniques, and the output of the Estimate Activity Duration process group is summarized below in Table 2-1.

The Estimate Activity Durations process involves the extraction of the information, activities, resources, and any other relevant inputs listed in Table 2-1 and using either one or a combination of tools and techniques to produces activity duration estimates (Sanghera, 2019).

### 2.3.2.1 Duration estimation tools and techniques

As shown below in Table 2-1, PMBOK 6th edition provides eight tools and techniques for estimating activity duration. This tools and techniques can be adopted in activity, phase or project duration estimation either in combination of one or more of them.

**Table 2-1 Estimate Activity Durations Process adopted from (Sanghera, 2019) and (PMI, 2017)**

<b>Inputs</b>	<b>Tools and Techniques</b>	<b>Outputs</b>
1. Project Management Plan:	1. Analogous estimating	1. Activity duration
• Schedule management plan	2. Parametric estimating	estimates
• Scope baseline: scope statement, WBS, WBS dictionary	3. Three-point estimating	2. Basis of estimates
2. Project documents:	4. Bottom-Up estimating	3. Updates:
• Activity list	5. Data Analysis	• Activity attributes
• Activity attributes	. Alternative analysis	• Assumptions log
• Milestone list	. Reserve analysis	• Lessons learned
• Assumptions log	6. Decision making	register
• Risk register	7. Expert judgment	
• Team assignments	8. Meetings	
• Resource calendars		
• Resource requirements		
• Resource breakdown structure		
3. Enterprise environmental factors		
4. Organizational project assets		

1) **Analogous estimating**; is a technique for estimating the duration or cost of an activity or a project using historical data from a similar activity or project.

- 2) **Parametric estimating**; an estimating technique in which an algorithm is used to calculate cost or duration based on historical data and project parameters.
- 3) **Three-point estimating**; a technique used to improve the accuracy of a single time estimate by using three activity durations; favorable (optimistic), unfavorable (pessimistic), and the most likely duration.
- 4) **Bottom-UP estimating**; the aggregation of detailed durations estimated into total quantity for each of the activity's durations.
- 5) **Data Analysis**; two of such techniques include Alternative analysis; where comparison of various levels of resource usage conducted to arrive at suitable duration and **reserve analysis** to investigate the amount of contingency required for the uncertainty involved in an activity duration estimates.
- 6) **Decision making**; several alternative duration estimates may come to final decision and voting can be adopted in such a scenario.
- 7) **Expert judgment**; is defined as judgment provided based upon expertise in an application area, Knowledge Area, discipline, industry, etc., as appropriate for the activity being performed.
- 8) **Meetings**; project teams may consider conducting office or site meetings to decide the appropriate activity duration.

From the eight tools and techniques recognized for duration estimation by PMBOK, this study used parametric duration estimation using linear regression. Parametric duration estimation attempts to develop a mathematical model by incorporating significant duration influencing factors. Identification of duration influencing factors is therefore the major and first step in developing parametric models. The following section discuss findings of previous studies on building duration influencing factors.

## 2.4 Building Construction Duration Influencing Factors

Identification of important construction duration influencing factors enables to develop a more functional model for future estimation. This section reviews the various factors that influenced the construction duration of building projects. Gab-Allah, et al., (2015)

indicated that identification of these factors can enable accurate assessment of the required project duration. While Nkado, (1995) indicated that the most important factors are those that can be identified from a project information and their influence can be modeled using mathematical methods together with judgmental analysis. The very basic use of identification of influencing factors is; to define the magnitude of their effect and develop a mathematical model that uses the relationship that exist between such influencing factors and construction duration.

Factors that contribute to the variation of both an activity and the project duration in general are reviewed. There are a number of risk factors that might cause variation in activity duration, and their influence can differ from one activity to another (Dawood, 1998). When considering the overall project duration variations, several researches used different categorization of individual factors, for the purpose of this paper it was found useful to classify individual set of factors in to two broad categories namely: 1) project related factors and 2) non-project related factors.

Project related factors that influence construction duration of building projects, are those that are associated with the nature and characteristics of projects, while non-project related factors are outside the control of the project.

Project related factors are either the technical parameters that describe the physical characteristics and size of the building or factors associated with the project stakeholders and their specific organization of the project.

Several studies used the term project scope to define the technical parameters that describe the physical characteristics and size of the building. The project scope term as used in literatures regarding identification of duration influencing factors here is different from the one used by PMI which defines it as; *“The work performed to deliver a product, service, or result with the specified features and functions.”* (PMI, 2017). Chan and Kumaraswamy, (2002) listed six influencing factors under the project scope category; Cost, Gross floor area (GFA), number of storeys, building type, contract procurement systems, and variations. The factors GFA, number of storeys are measures of building physical size which are described as technical parameters category together with building height and volume of work (Gab-Allah, et al., 2015). Number of basement level, (Elhag & Boussabaine, 1999; Lim, et al., 2017 )

The construction cost which is the budget required to construct a facility for intended purpose was the first project scope parameter identified to affect the duration of building projects by Bromilow, (1969) in Australia, who developed a mathematical relationship that relate cost and construction time. Bromilow, (1969) discussed the importance of using cost as a measure of the physical size of the project. Cost also considered as an influencing factor by (Chan & Kumaraswamy, 1995; Gab-Allah, et al., 2015; Guerrero, 2016). Chan and Kumaraswamy, (1996b) identified 84 variables which were further grouped under seven major sections: (A) general organization information; (B) project characteristics; (C) client characteristics; (D) architect/engineer; (E) building contractor; (F) general assessment; and (G) construction cost/time information.

Apart from project scope factors the following project related factors had been identified as duration influencing ones such as: building type, client type, and design or project complexity. Project complexity tend to create difficulty to both the contractor and the design team which in turn put a major influence on the construction duration of a building project. This factor was mentioned in a number of previous studies; e.g. (Guerrero, et al., 2014). Walker, (1995) identified three construction time performance affecting factors which are related with the project management team communication these are: construction management effectiveness, the sophistication of the client and the client's representative in terms of creating and maintaining positive project team relationships with the construction management and design team, design team effectiveness in communicating with construction management and client's representative teams (Walker, 1995).

The influence of the type of client within building works on the duration of a building project was determined by; (Kaka & Prince, 1991; Nkado, 1992; Chan & Kumaraswamy, 1996b). To determine the influence of type of client on project duration, researchers adopted to classification of collected sample data into their respective client (e.g. private, government) and they have identified a variation on duration for the different classes used. A detailed discussion of the models presented in the next section under modeling construction duration. A study in Nigeria that used 96 building project samples for modeling the relationship between time influencing factors identified the following key factors which are here we categorized them under project related factors; delay in progress payment by owner, lateness in revising and approving design document by owner, delay

in delivering the site to the contractor by the owner, change order by owner during construction, poor site management and supervision by contractors (Ahmadu, et al., 2015).

Table 2-2 below provide a short list of influencing factors considered as project related ones together with those researchers who identified or considered them as one contributor in a various categorization of factors.

**Table 2-2 Project related factors that affect duration of construction building projects**

	Researchers cited					
	(Thornton, 1988 )	(Chan & Kumaraswamy, 1995)	(Nkado, 1995)	(Dissanayaka & Kumaraswamy, 1999 )	(Elhag & Boussabaine, 1999)	(Gab-Allah, et al., 2015)
Area (size)	✓	✓			✓	✓
Building type	✓	✓	✓		✓	✓
Complexity		✓	✓	✓	✓	✓
Construction method	✓				✓	
Cost/value		✓				
Height					✓	✓
No. of floors		✓				✓
No. of basement floors					✓	
Quality of finishing					✓	✓
Quantity of work						✓
Site conditions					✓	✓
Type of cladding & external wall					✓	
Type of foundations					✓	
Type of structures					✓	
Type of contract	✓	✓				
Change orders		✓		✓		

Further apart from those factors discussed; material and equipment failure, incomplete design scope, defective design, design changes, fluctuation in labour productivity, artificial

obstruction, and subcontractors default were also suggested as influencing ones (Dawood, 1998).

Non-project related factors are those factors for which the project team has low or none control over them, further these factors tend to be much uncertain as compared with their project related counter parts. Several factors affecting the time performance of the construction stage cannot be accurately predicted (e.g., weather, unforeseen site conditions) or the required information is not available in the early stages of the project development (Sousa, et al., 2014).

The project environment factor category comprising of; physical, economic, socio-political, and industrial relations by Chan and Kumaraswamy, (2002) fall under non-project related factors category of this study. Further factor category by the name external factors identified by (Nkado, 1995; Elhag & Boussabaine, 1999) and market conditions described by (Elhag & Boussabaine, 1999) also can be taken as non-project factors. Type of soil and site condition; Weather conditions;

Ahmadu, et al., (2015) identified the effect of rain on construction duration which is included under the effect of weather, was also identified by researchers who applied either probabilistic method or a fuzzy logic to quantify the influence; e.g (Boussabaine, 2001; Kumar & Reddy, 2005; Rezakhani & Maghiar, 2019). In addition to the influence of raining, average temperature and humidity were considered as influencing the construction duration of building specifically the concrete pouring activity (Golizadeh, et al., 2015).

Project location also can be considered as a non-project related factor, and it was found to be a main influencing factor specially in the following studies; (Dursun & Stoy, 2011a; Dursun & Stoy, 2011b)

It is clear that an earlier detection and understanding of the influence of non-project related factors enable to derive a construction method that allow accomplishment of target project duration. But due to the very nature of their uncertainty it is difficult to incorporate them in duration estimation models. The manager must be aware of the effects uncertainty may have so that he may make a rational decision as to what level of risk to accept in the light of the circumstances at the time (Bennett & Ormerod, 1984). there are a number of risk factors that might cause variation in activity duration, and their influence can differ from one activity to another.

## 2.5 Modeling Construction Duration of Building Projects

The concept of modeling as used here in this thesis identifies the process of developing an empirical or mathematical model. While the term model refers to any mathematical or empirical relationship that is developed using any selected approach (e.g. linear regression) to show the influence any of construction duration influencing factors.

This section is separated in to two major subsections; the first addressed the linear regression modeling approach for construction duration which is much wider than the second section which focus on other modeling approaches. The first section is more detailed than the second this is done on purpose. This is because this thesis adopts linear regression approach to model the construction duration of structural framework, therefore in order to gather essential foundational concepts from the previous studies that adopted linear regression modeling. While under the second class of modeling approach only certain number of researches are developed and the strength and weakness of such an approach is presented. Modeling approaches like Fuzzy logic, Artificial Neural Network (ANN), and simulation are presented.

### 2.5.1 Linear Regression Approach for Modeling Construction Duration

As seen in the previous section the PMBOK identified parametric estimating method among the other tools and techniques for duration estimation. In this section of the literature review parametric duration estimation models developed using statistical regression as applied to building projects are discussed in depth.

Fahrmeir, et al., (2013) explained regression as the most popular and commonly used statistical methodology for analyzing empirical problems in multi discipline areas; social sciences, economics, engineering, and life sciences. Correspondingly, there exist a large variety of models and inferential tools, ranging from conventional linear models to modern non- and semiparametric regression.

It would be helpful to define certain terminologies first in this section that will be used throughout the literature review and the whole paper.

- **Terminologies and abbreviations used**

*Variable*; any influencing factor or the quantity that can assume any set of numerical value. It can be; independent, dependent, qualitative, or quantitative as defined below.

*Dependent/ response variable*; is a variable of interest for which we want to find a model showing the interrelationship and influence of other variables. For this particular case construction duration or time is the response variable.

*Independent/ predictor variable*; is a variable that is selected either based on previous study or using the judgement of the researcher assumed to influence or to have a relationship between the response variable.

*Quantitative variable*; a variable that can be quantified using a numerical value. E.g. the floor area of a building.

*Qualitative (categorical) variable*; a variable for which a numerical value cannot be associated with e.g. project location.

*Dummy variable*; a variable used to include qualitative variables into model but the value doesn't signify the magnitude rather indicate the variable of interest; e.g. building type defined by dummy variable 1 for commercial building and 2 for residential building.

*Model*; any empirical model that showed a relationship between construction duration or time and any number of influencing factors or predictor variables.

*Modeling*; any approach that is used to develop an empirical model of interest.

*Linear regression analysis*; a statistical regression approach that attempt to fit a number of observation data set to a straight line of the form  $y=mx + b$ . A linear relationship is first assumed to exist among the variables of interest.

*Simple linear regression (SLR)*: considers the influence or the linear relationship that exist between a single predictor variable  $x$  and a dependent variable  $Y$  the relationship is defined in equation 1 below, where the intercept  $\beta_0$  and the slope  $\beta_1$  are unknown regression coefficients and  $\varepsilon$  is the error term between the predicted and actual data commonly known as residual;

$$Y = \beta_0 + \beta_1 x + \varepsilon \quad (1)$$

*Multiple linear regression (MLR)* model, the dependent variable or response is related to  $k$  independent or predictor variables. The model is indicated in equation 2 below;

$$Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_kx_k + \varepsilon \quad (2)$$

*Coefficient of determination ( $R^2$ ):* a statistical metric having a value in the range of 0 and 1 which is reported as a percentage used to measure how much of the variability in the response variable is explained by a given model. the closer the  $R^2$  value to 1 the more variations or accurate predictions can be explained.

The terminologies defined above are based on their use on this paper and may hold other definitions to other studies.

Modeling of construction duration concept goes some 50 years back, where F. J. Bromilow in Australia derived a mathematical relationship between the construction time and construction cost of building projects. His research grabbed the attention of several researchers to study and validate the model for their respective country and for both building and infrastructure projects. Later the model gained acceptance and known to be Bromilow's Time-Cost model (BTC) or Time –cost relationship (TCR), here in this paper the term BTC used to refer to this model.

BTC model is the focus point of this section of the review. First the original BTC model reviewed, later some selected validations to the BTC model reviewed, and finally three types of variations to the original model are discussed.

### ***2.5.1.1 Bromilow's Time-Cost (BTC) Model***

The most cited work of Bromilow which was printed in *Building Forum journal in September 1969* is considered as the first modeling attempt of construction time (duration). The Commonwealth Scientific and Industrial Research Organization (CSIRO), division of building research of Australia, in 1967 undertook a pilot study to investigate the performance of building contracts to identify the magnitude of departure from expected contract time (Bromilow, 1969). Project data of 329 building projects completed between July 1964 and June 1967 from three Australian cities were used to determine the deviation of expected contract time from the actual construction time. The result of the pilot study showed that the overall average extra time taken for building projects exceeded 40 percent.

Clients as observed by Bromilow, (1969) expected their projects to be completed on contract stipulated time even though they agree on legal conditions for extension of time

(Bromilow, 1969). Much of construction clients' dissatisfaction is related with failure of the industry to meet cost and time expectations. Bromilow thought that the high deviations of expectations and reality might have come from optimistic contract time. From data of 309 projects actual and construction time it was found that only 37 projects completed on or before time.

Bromilow attempted to derive a time performance standard by analyzing the construction times actually being taken by 303 building projects. The construction time was found to be highly correlated with the size of the building as measured using actual construction cost and this relationship is given in equation 3 (Bromilow, 1969) .

$$T = K \cdot C^B \quad (3)$$

Where: T – duration of construction period from date of site possession to practical completion, in working days,

C – final cost of project in millions of dollars, adjusted to constant labour and material prices,

K – constant describing the general level of time performance for a \$1million project,

B – constant describing how the time performance is affected by the project size as measured by project cost.

The numerical values of the constants K and B become 350 and 0.3 respectively. Which is interpreted as a project worth of \$1 million is expected to take 350 working days for completion (Bromilow, 1969).

Bromilow expressed that cost is useful measure of project complexity, quality and physical size (Bromilow, 1969).

Equation 3 in the later years has come to be known as Bromilow's Time Cost Model or BTC model, the remainder of the literature uses this short term to refer to this particular relationship.

### ***2.5.1.2 Validations for BTC Model***

After the discovery of the relationship between construction time and project size as measured with cost, several studies had been conducted, with the aim of validating the

relationship and setting a standard performance tool for the construction industry in a given country and still researches are being done year after year. Here in this section notable researches are selected and presented in detail further a table of collection of similar researches is provided at the end of this section to show the trend in the application of the BTC model worldwide.

BTC model had been applied in various types of civil works (road, sewerage, rail way, dam) but this study focuses on studies conducted with regard to building construction only except for some exceptional cases.

A comprehensive review of literature spanning through the years 1969 to 2016 focusing on modeling construction time indicated that Bromilow further confirmed his developed model in 1974 and 1980, thus from the period of 1969 to 1980, Bromilow was the only researcher who had conducted studies on modelling construction duration (Mensah, et al., 2017)

A) (Bromilow, et al., 1980 )

After C.S.I.R.O conducted the pilot study in 1967 and Bromilow derived the relationship between time and cost as in equation 3 , the division of building research required the collaboration of the industry stakeholders to monitor the time performance of building projects (Bromilow, et al., 1980 ). The Australian Institute of Quantity Surveyors (AIQS) conducted a questionnaire survey to gather information relating to building projects completed during the period of 1970 to 1976.

The AIQS survey was able to gather 419 projects; 290 government and 129 private projects. 13 government and 11 private were excluded due to high abnormalities, which resulted in a total sample of 395 projects.

The survey conducted to investigate the deviation in time taken for construction from the previous time performance of the 1969 Bromilow study. The result of the average time taken for a \$1 million project as described in the constant coefficient K of equation 3 is tabulated below in Table 2-3.

**Table 2-3 AIQS survey values of K adopted from (Bromilow, et al., 1980 )**

	Working days K		Change from 1969-survey Percent (%)	
	Private	Government	Private	Government
Pre-1974 projects	232	335	+7.4	+14.3
Post-1974 projects	243	406	+12.5	+38.6
Total	238	358	+10.2	+22.2
Pre-1969 projects	216	293	-	-

It had been clear that the time taken for projects increased approximately between 7% and 40 %. Bromilow, et al., (1980 ) deduced that this increase in the actual time taken by projects were due to the over heated economy of Australia in 1973 characterized by shortage of materials, skilled labour and managerial expertise and extensive delays arising from disputes in labour matters.

**B) (Ireland, 1983)**

Ireland, (1983) conducted a PHD dissertation work on analysis of the effects of managerial actions on the objectives of reducing time, reducing cost and increasing quality of high rise building constructions in Australia. As part of the disertaion study, Ireland, (1983) validated the BTC model using data of 25 high rise office building projects. The model was able to describe 58% variation in construction time and is given below;

$$T = 219C^{0.47} \tag{4}$$

where T is the construction time measured in days (excluding days when no work was done, such as Sundays and public holidays, but including days on which strikes were held) and C is the construction cost in millions indexed to June 1979.

Ireland, (1983) further indicated that an increase in construction time resulted from; an increase in variations to the contract, the complexity of the building, the number of storeys and industrial disputes.

**C) (Kaka & Prince, 1991)**

Kaka and Prince, (1991) indicated that a strong relationship exists between the cost and duration of construction projects in United Kingdom (UK). They used a sample of 661 building projects, which included all types of commercial, industrial, residential and public projects. And a sample of 140 road contracts. Kaka and Prince, (1991) used the following catagorie: type of client; (public, private); type of project (Building, civil); type of tender

(open competition, selected, negotiation); form of tender (fixed price, fixed adjusted) and developed time-cost relationship models for both samples under each category. The results indicated a much stronger relationship which was expressed in Pearson’s correlation coefficient of the range 49% to 97%.

**D) (Chan & Kumaraswamy, 1995)**

Chan and Kumaraswamy in Hong Kong conducted successive studies to determine both construction duration influencing factors and standard time performance model. Their work investigated the relationships between different project characteristic variables such as the construction duration, construction cost, total gross floor area and the number of storeys in the case of buildings (Chan & Kumaraswamy, 1995). They were able to collect a sample of 111 projects from a questionnaire survey and collected sample classified into three different categories: government buildings, private buildings and civil engineering works selected model results are presented in table.

**Table 2-4 Time-cost performance model for selected projects in Hong Kong (Chan & Kumaraswamy, 1995)**

Type of project	Actual construction time		
	K	B	R
Total public buildings	216.3	0.253	0.79
Total private buildings	250.9	0.215	0.65
Total civil project	291.4	0.205	0.78

**E) (Ng, et al., 2001)**

Ng, et al., (2001) tested and refitted BTC model with a new set of data for Australian construction projects completed between 1991 and 1998. A sample containing 93 public and private building projects used. Samples were categorized using; sector (public, private), contractor selection (selective, open, negotiation), project type (recreational, industrial, educational, other), contract (lump sum, design & construct, construction management, other).

$$T_{public} = 129C^{0.32} \tag{5}$$

$$T_{private} = 132C^{0.3} \tag{6}$$

$$T_{Overall} = 131C^{0.31} \tag{7}$$

### F) Other selected BTC model validations

This section is not intended to list all works that validated the time-cost relationship, rather, it is to make point on how researchers applied it to various categorization and how the values of the constants of the model change from a given category to the other. Readers could find a number of journal articles from several databases and make understanding of other models too. A table consisting of some other works as a reference is provided below.

**Table 2-5 Selected examples of BTC model validations for building projects**

Author(s)	Country	Number of samples	BTC model
(Chan, 1999)	Hong Kong	110	$T_{all} = 152C^{0.29}$ $T_{public} = 166C^{0.28}$ $T_{private} = 120C^{0.34}$
(Chan, 2001)	Malaysia	51	$T = 269C^{0.32}$
(Choudhury & Rajan, 2004)	USA	55	$T = 18.96C^{0.39}$
(Chen & Huang, 2006)	Taiwan	132	$T = 18.085C^{0.194}$
(Le-Hoai, et al., 2009 )	Vietnam	77	$T_{all} = 93.6C^{0.338}$ $T_{public} = 98.1C^{0.343}$ $T_{private} = 87.2C^{0.348}$
(Odabaşı, 2009)	Turkey	7	$T = 161.74C^{0.58}$
(Jarkas, 2015)	Kuwait	187	$T_{residential} = 28.79C^{0.192}$ $T_{office} = 7.92C^{0.277}$
(Bayram, 2016)	Turkey	530	$T_{contract} = 192C^{0.467}$ $T_{actual} = 209C^{0.353}$

Before concluding this section where a number of validation researches reviewed, the question; *is the BTC model always valid?* Is asked. A number of researches proved that it is not always so. Ogunsemi and Jagboro, (2006) in Nigeria after analyzing a sample of 87 building projects where they used 70 of them for model development and the rest for validation purpose, conclude that BTC model not applicable for Nigerian construction (Ogunsemi & Jagboro, 2006). A similar conclusion was reached by Ameyaw, et al., (2012) in Ghana, where they assessed the applicability of the BTC model on 62 building projects. The coefficient of determination ( $R^2$ ), which is defined under the common terminologies

and abbreviations earlier was found low; 20.5% for (Ogunsemi & Jagboro, 2006), while 37.8% reported by (Ameyaw, et al., 2012). But contrastingly  $R^2$  value of 33.7% was reported to be acceptable by (Hoffman, et al., 2007).

Such contrasting results made researchers to re-test the BTC model to their respective country and to a different classis of cases, with these variations of BTC model which identified the gap in it and derived new approaches are discussed.

### ***2.5.1.3 Variations to BTC Model***

The BTC model which defines the mathematical relationship between construction time and the project cost as discussed in the previous section validated by different authors in different countries around the world, this subsection however shows the variations to the time-cost relationship. The variations to the model as witnessed in previous studies came from three aspects; 1) dependent variable variation, 2) predictor or independent variable variation, and 3) incorporating more influencing factors in a model. The first variation addressed the appropriateness of construction time as a dependent variable, the second questioned the use of cost as independent variable describing the physical size of building, while the last attempted to develop a model that includes the effect of various influencing factors in a single model. a brief discussion is followed below for each variation cases recognized in literature.

#### ***2.5.1.3.1 Construction time or construction speed?***

A number of researches reviewed was identified to vary from the original BTC model in the choice of construction speed as a dependent variable instead of the construction time as used in the original BTC and in validated successive studies.

The average progress of construction over the construction duration defines the construction speed. Normal units of measure are revenue (€ revenue/month), the area completed ( $m^2$  gross external floor area/month) or volumes completed ( $m^3$  gross volume/month) (Stoy, et al., 2007).

Stoy, et al., (2007) used a data of 115 residential building projects constructed in Germany that were accessed from a single database to develop a construction speed model. The logarithmic form of construction speed, measured in  $m^2$  gross external floor area/month merges as the appropriate parameter for the study (dependent variable). The derived model is presented below in equation 8;

$$\ln(y) = 4.753 + 0.0002X_1 - 0.001X_2 \quad (8)$$

where:

$y$  - is Construction speed (m<sup>2</sup> gross external floor area/month).

$X_1$  - Absolute size (m<sup>2</sup> gross external floor area).

$X_2$  - Project standard (€ building construction costs/m<sup>2</sup> gross external floor area).

The model thus identified that project size, which is measured in m<sup>2</sup> gross external floor area, and the project standard, which is measured in € building construction cost/m<sup>2</sup> gross external floor area, as significant drivers of construction speed (Stoy, et al., 2007).

Guerrero, (2016) in Spain, in his doctoral dissertation work also used construction speed as a dependent variable for measuring the construction duration. A hypothesis formulated as “*Construction speed is a more appropriate dependent variable than construction time in order to generate predictive models for estimating the duration of the construction process of new builds.*” Was validated against the null hypothesis using the Pearson's correlation coefficient (R) statistic. When construction time is used as a response variable, the analysis results showed weak correlation values with the independent variables. On the contrary, using the construction speed as a dependent variable very high correlation values were obtained. A simple linear regression was conducted further to reinforce the findings of the Pearson's correlation and the values of adjusted coefficient of determination (adjR<sup>2</sup>) were used to compare the resulting models from using both construction time and speed as dependent variable. Construction speed variable resulted in a adjR<sup>2</sup> values much higher than its' construction time counterpart (Guerrero, 2016). The developed construction speed model is shown below in equation 9.

$$\ln(\text{speed}) = -0.609 + 0.928\ln(T\_GFA) - 0.167\ln(T\_Floor) - 0.232\ln(\text{standard}) \quad (9)$$

Where; *Speed* is the construction speed, *T\_GFA* is the total gross floor area, *T\_Floors* is the total number of floors and *Standard* is the total cost of construction (*T\_Cost*) divided by *T\_GFA*.

#### 2.5.1.3.2 Cost or physical size (gross floor area & Number of floors)?

The pioneering research by Bromilow, (1969) which is discussed in much detail initiated several researches in the civil engineering domain to evaluate if such a relationship between construction time and cost existed in a given type of construction (Building,

highway, railway, dam). However, the appropriateness in the use of cost as a measure of the size of a project came to be the center of attention in the building construction projects. Variations of BTC that questioned and argued the use of cost as a predictor variable are carried out by number of researchers.

Khosrowshahi and Kaka, (1996) after indicating that project total cost and duration to be influenced by a number of variables which act individually and in combination, commented that the extent of the effect of these variables is dependent on the nature of the project. For building projects which are the focus of this thesis a clear argument raised from different authors on the use of cost as predictor variable. The use of physical size indicator variables of building; namely gross floor area (GFA) and number of floors suggested by researchers.

The influence of GFA and number of floors on construction duration of building were examined using a similar model as the original BTC model the equations are given below (Chan & Kumaraswamy, 1995);

$$T = LA^M \tag{10}$$

$$T = FS^G \tag{11}$$

Where; T- is the construction time (days), A –is the GFA (m<sup>2</sup>), S –is the number of storeys in a single block, while the coefficients L&F, and F&G correspond to the constants K and B of the BTC model respectively.

Chan and Kumaraswamy, (1995) were able to find a relevant Pearson’s correlation between GFA and construction time of for government and private building projects in Hong Kong, the minimum being 0.65 and maximum 0.79. Also, similar correlations in the range of 0.6 up to 0.78 were observed between time and number of storeys of building.

Love, et al., (2005) after conducting multiple regression analysis using 161 construction projects that were completed in various Australian states, they concluded that GFA and the number of stories in a building are key determinants of time performance in projects. Furthermore, the results indicate that cost is a poor predictor of time performance (Love, et al., 2005). Further also Love, et al., (2005) criticised the use of time-cost relationship

developed in BTC model by indicating that cost is not known before the completion of a project, also project cost tend to vary from what was estimated during the planning phase. The model recommended for prediction of time based on GFA and the number of floor levels resulted in a higher adjusted R-squared value of 0.96 and is presented below;

$$\log(T) = 3.178 + 0.274\log(GFA) + 0.142\log(Floor) \quad (12)$$

A contrasting conclusion was provided by Bayram, (2016), in Turkey after analyzing 530 Turkish public building projects. He used 80% of the total data (i.e. from 424 projects) to establish the models while the remaining 20% (106 projects) were used for validation of both BTC model and the previously presented Love, et al's., (2005) time floor relationship he concluded that BTC model is a superior over the time-floor relationship. Cost emerged as a more significant predictor of construction time (Bayram, 2016).

Guerrero, et al., (2014) on the other hand indicated GFA has greater influence than cost on project duration but both factors are necessary to achieve a forecast model with the highest accuracy. Similar conclusions also provided by; (Chen & Huang, 2006) for school reconstruction duration model in Taiwan and (Choudhury, 2012) for construction time of educational projects in Texas.

Choudhury, (2012) used a data of 39 educational building in Texas and measured the hypothesized time influencing variables; cost, delivery method and GFA. He found that a significant relationship existed between time and GFA, and the model was able to describe around 73% variation in construction time. It is presented below;

$$Time = 0.067GFA^{0.454} \quad (13)$$

Chen and Huang, (2006) used data for 132 school reconstruction projects in central Taiwan, which received the most serious damage from the Chi-Chi Earthquake in 1999, were collected and analyzed. Both cost and GFA showed a relatively same correlation with construction time and were used in simple linear regression analysis and coefficient of determination R<sup>2</sup> was used to select a model that covers a larger amount of variation in time variable. The cubic form of regression using GFA as predictor variable described

around 75 % of the variation in construction time, which was higher than described by its' cost counterpart. The model is presented below;

$$T = 131 + 0.018A - 1.085E - 6A^2 + 3.675E - 11A^3 \quad (14)$$

Where; T denotes construction time in days and A, the floor area in m<sup>2</sup>.

To conclude this section, it is clear that there is a disagreement among researchers on which predictor variable to use; i.e. either cost or physical characteristics described by GFA or number of storeys. As discussed above Love, et al., (2005) put a convincing argument on the choice between either cost or GFA and prioritizes GFA over cost. In practice, the GFA and number of floor levels are known before a project commences while cost is not (Love, et al., 2005). Since this thesis dedicated to derive a model that will be utilized during the early planning phase of a building project, it was found appropriate to use GFA instead of cost. Further the lack of proper cost and project information documentation practice by local contractors made the research to leave cost from hypothesized list of influencing factors. Moreover, also the time effect of money which is the main concern while using BTC model came as a main barrier to derive such a relationship for the current case. This is because the lack of usable cost index like tender index or building cost index, from any statistical agency in Ethiopia convinced to left cost out.

#### 2.5.1.3.3 Inclusion of more duration influencing factors

The third variation to the BTC model came from a major criticism of the model for only considering the influence of construction cost, while it was evidenced by many researchers that construction time is influenced by several factors. For this criticism researchers around the world attempted to come up with a variation of BTC by including other relevant influencing factors using multiple linear regression (MLR) technique. Here we review the important uses and findings from the technique.

28 journals concerned only on modeling of construction duration of building using linear regression were reviewed, out of which 73% used MLR technique (50% only MLR while the rest 23% both SLR and MLR). This shows that MLR had been used much as compared to SLR.

MLR enables the consideration of a large set of construction duration influencing factors, but as discussed in previous section simple linear regression (SLR) only considers the

effect of a single variable. The necessity for researchers favoring MLR came from two aspects; either to increase the variability described in a model (higher  $R^2$  value) or inclusion of qualitative variables (those for which a numerical value is not available) in a model.

From inclusion of a larger set of influencing factors the work by (Chan & Kumaraswamy, 1999b) measured the effect of 84 variables while (Nkado, 1992) attempted to value the effects of 62 construction time influencing factors. Three methods are used to select the most influential factors among such a large number of hypothesized variables; Forward selection, Backward elimination, and Stepwise method.

Forward selection technique starts with an equation containing no predictor variable, i.e. with the constant term of the regression equation, then adds the variable with a higher correlation with the response variable then check if the coefficient is significantly different from zero if so retained in the equation or dropped, this iterative process terminates when the last variable entering the equation has an insignificant regression coefficient or if all variables are included when found significant (Chatterjee & Hadi, 2006).

The Backward elimination technique opposite to the forward selection starts with all the predictor variables in the equation and attempt to drop a variable when an elimination criterion is fulfilled (variables contribution to the reduction of error sum of squares used as criteria). The stepwise method is a modification of the forward selection method where a variable that entered during earlier stage of selection may be eliminated at a later stage (Chatterjee & Hadi, 2006). The stepwise selection is the most used for modeling duration influencing factors e.g. (Chan & Kumaraswamy, 1996a; Chan & Kumaraswamy, 1996b; Choudhury, 2012; Guerrero, et al., 2014). Here in this thesis only a short description of the techniques is provided in order to familiarize the terms with the reader, but readers can refer a complete regression books like (Chatterjee & Hadi, 2006) or an introductory statistics books like (Montgomery, et al., 2011). In this thesis all the three methods were used.

Construction duration influencing factors by their nature are either quantitative (measured in numerical value) or qualitative (not quantified in numerical values). Further also as discussed in the previous section most duration influencing factors tend to be subjective to the response of the person, which might change from time to time. It was indicated that

project complexity as a main influencer but how one evaluates the term varies from individual to individual. MLR provides two options for the inclusion of the effect of qualitative variables in to predictive models; either developing a separate model by separating the data sample into classes of qualitative factors or using a dummy variable that quantifies the qualitative variable of interest.

Qualitative variables also known as categorical variables can be included into a MLR model using a dummy variable coding technique which is defined for a variable  $X$  containing  $C$  categories,  $X \in \{1, \dots, C\}$ ; the variable  $X$  can be coded with  $C-1$  dummy variables  $X_1, \dots, X_{C-1}$ ;  $X_j = \begin{cases} 1 & \text{if } x = j \\ 0 & \text{otherwise} \end{cases} j = 1, \dots, C - 1$  (Fahrmeir, et al., 2013).

Nkado, (1992) used MLR to model the effects of the following qualitative variables on construction duration of 29 commercial buildings; end use, cladding type, presence of atrium, building location, intensity of services and site accessibility together with other quantitative variables. While the influence of project type, procurement method, and tender type was evaluated by (Love, et al., 2005). Lin, et al., (2011) in Taiwan using 56 project data identified the effect of number of change orders and rainy days. The study by Chan and Kumaraswamy, (1996b) that used a project data of 56 "standard 'Harmony' type domestic blocks of the Hong Kong Housing Authority resulted in the identification of three important qualitative factors for duration forecasting: type of foundations; information flows between architect/engineer and contractor; presence/absence of precast facades; and type of scheme (rental/ purchase). While the categorical factors ground conditions for construction and labour productivity were identified by (Chan & Kumaraswamy, 1996a) from analysis of 15 standard (New Cruciform) type housing blocks using MLR. A model consisting the influence of client sector, contractor selection method, contractual arrangements, and project type, for forecasting actual construction time were also developed by (Skitmore & Ng, 2003).

The other main benefit of modeling the relationships that exist in construction duration influencing factors is the ability to describe a large amount of variability in duration. The variability observed from a single model measured by either the coefficient of determination  $R^2$  or adjusted  $R^2$  ( $\text{adj}R^2$ ) observed to be higher than the one obtained using SLR.  $R^2$  value in the range of 74% to 98% was observed in literatures that used MLR technique. 74.6 % variation was explained by (Lim, et al., 2017 ), (Lin, et al., 2011) model

explained 92%, and (Chan & Kumaraswamy, 1996a) were able to achieve a model explaining 98 % variation in construction time. An  $\text{adjR}^2$  value of 96% was observed from (Love, et al., 2005), and 92.7% in (Khosrowshahi & Kaka, 1996) study. Here it is not concluded that such high  $R^2$  and  $\text{adjR}^2$  values are obtained by using only MLR method rather a larger amount of variability was able to be covered by inclusion of more influencing factors into the model. But a number of studies that attempted for selecting a best performing model among a set of SLR and MLR models found the MLR model to outperform its' SLR counterpart e.g. (Odabaşı, 2009).

From this section discussion it is understandable that, MLR technique is useful variation of the original BTC model. The importance came in two-fold; the first crating ability to include the effect of qualitative (categorical) variables in to models using dummy variable and the second increasing the amount of variation explained by a model. The main objective of a given forecasting or predicting model is explaining a larger portion of variability in the response variable. Therefore, here also this study utilized both mentioned benefits of MLR technique.

### **2.5.2 Other approaches for modeling construction duration**

In the previous section a closer and detailed examination into linear regression modeling approach of construction duration forwarded. But, the approach came with shortfalls where researchers show interest to use other approaches of modern mathematics, simulation, and Artificial Intelligence (AI). Vahdani, et al., (2016) explained that time data applications in the construction industry are complex and nonlinear in nature, which is opposite to the major assumption behind linear regression models. another criticism is its' highly sensitive to outliers, and is generally considered unsuitable to account for the large number of variables and the interactions among them (Jin, et al., 2016).

This section reviewed three other approaches utilized in previous studies; Fuzzy logic, Artificial Neural Network (ANN), and Simulation. Their strength and weakness for the application of duration prediction models addressed.

#### **2.5.2.1 Fuzzy Logic Based Models**

Activity duration estimation involves subjectivity, vagueness, or imprecision which affect the accuracy of an estimate (Ayyub & Haldar, 1984; Kumar & Reddy, 2005; Pawan & Lorterapong, 2016). This subjectivity and vagueness arise from the fact that some of the

factors that influence an activity and project duration are qualitative, for which assigning numerical value is difficult task. (Zadeh, 1965) as cited in Zimmermann, (2010) introduced the concept of Fuzzy set to tackle this problem (Zimmermann, 2010). Fuzzy set is used to convert a linguistic variable which is defined as a “*variable, the values of which are words, phrases, or sentences in a given language.*” (Ayyub & Haldar, 1984) into a numerical value. An example of a linguistic variable can be, labor experience where its values can be expressed as, "long experience" or "short experience,". Such variables are meaningful in real world but the incorporation of such variables in mathematical models was not clear until the concept of fuzzy set developed by Zadeh.

Fuzzy is considered as a branch of modern mathematics that help modeling vagueness and solve complex real world problems, such complexity of a problem might come from incomplete previous data (Chan, et al., 2009). Chan, et al., (2009) indicated four application areas of fuzzy in the construction management domain: decision making; performance; evaluation/assessment; and modeling. The estimation of both time and cost performance of a project had been conducted using either fuzzy set/fuzzy logic and a hybrid fuzzy technique. Hybrid fuzzy technique uses a combination of fuzzy technique and neural network to enhance the performance of the model.

Fuzzy set/fuzzy logic for incorporating qualitative factors in both activity and project duration estimation arose the interest of researchers (e.g. (Ayyub & Haldar, 1984; Kumar & Reddy, 2005; Maravas & Pantouvakis, 2011; Pawan & Lorterapong, 2016)). The uncertainty in duration estimation where linguistic variables are encountered in most scenarios had been solved by the application of the fuzzy set theory. Boussabaine, (2001), and Vahdani, et al., (2016) used a Neurofuzzy technique for estimating project duration. Neurofuzzy technique is a combination of a fuzzy and neural network technique. It consists of the ability of fuzzy technique to represent qualitative, vague and imprecise concepts and the learning power of neural network (Boussabaine, 2001).

#### **2.5.2.2 Artificial Neural Network (ANN) Based Models**

The complexity of real-world system (process) to be presented as simpler yet informative model is a continual interest to various industry sectors. The construction industry is one which seeks such interventions to enhance the decision-making process involved. Artificial Neural Network (ANN) which resembles the human brain gained much attention to solve complex nonlinear problems in the construction sector (Guerrero, 2016).

Bhokha and Ogunlana, (1999) defined ANN as a branch of artificial intelligence (AI) whose structures are based on the biological nervous system. The efficient learning ability of ANN from previous data and generalizing to new data cited as one major characteristic similarity with the human brain (Bhokha & Ogunlana, 1999 ). Further ANN can provide meaningful answers even when the data to be processed include errors or are incomplete ( (Bhokha & Ogunlana, 1999 ; Petrusseva, et al., 2012; Gab-Allah, et al., 2015; Golizadeh, et al., 2015). ANN like the human nervous system composed of interconnected neurons forming a network. The components of ANN are (Guerrero, 2016):

- A set of inputs  $X_j$
- A set of Synaptic weights  $W_{kj}$ , affecting each neuron input.
- An adder for summing up the inputs, weighted by the respective synapses of the neuron.
- An external applied bias  $b_k$ , which has the effect of increasing or lowering the net input of the activation function, An activation function  $f$  that limits the amplitude of the output of the neuron.
- An output value  $Y_k$

Figure 2-3 below shows a graphical illustration of the basic components of ANN.

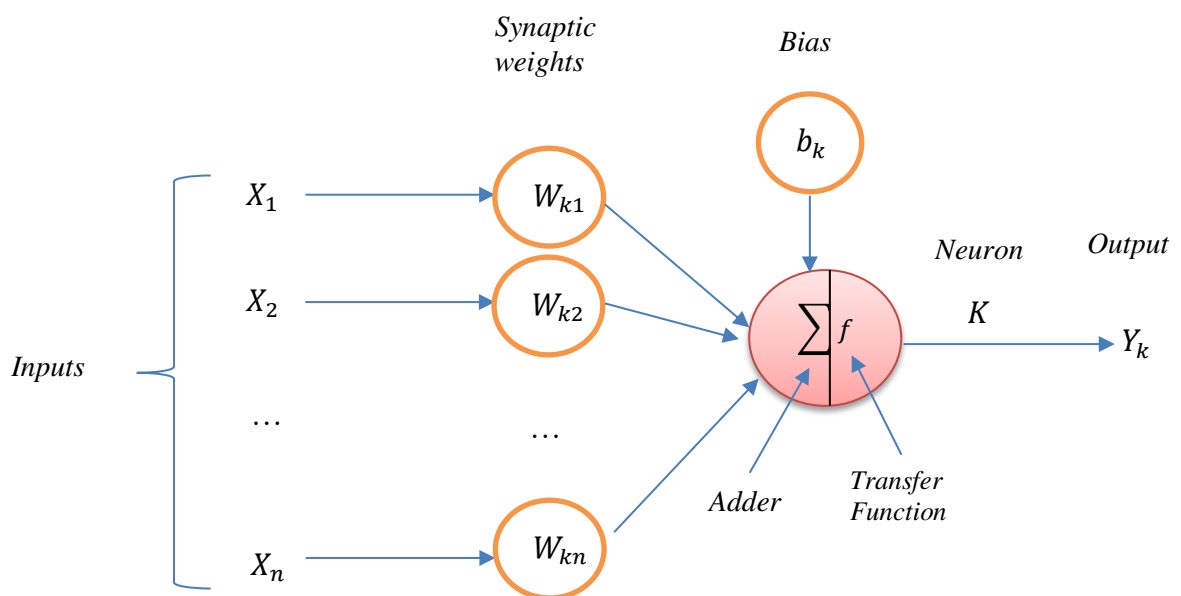


Figure 2-3 Operating diagram of an Artificial Neuron adopted from (Guerrero, 2016)

The mathematical description of the artificial neuron k was introduced by (Haykin, 1999) as cited in (Guerrero, 2016) as shown in equation 15.

$$Y_k = f \left( b_k + \sum_{j=1}^n W_{kj} X_j \right) \quad (15)$$

Where:  $Y_k$  is the output value;  $f$ , represents the transfer function;  $b_k$ , is the bias term;  $n$ , is the number of synaptic weights;  $W_{kj}$ , is the value of the synaptic weight  $j$ ; and  $X_j$  is the value of the input  $j$ .

Table 2-6 below provided by Maier and Dandy (2000) as cited in (Guerrero, 2016) contains glossary of the commonly used ANN terminologies and their statistical equivalents.

**Table 2-6 ANN terminologies and their statistical counterparts adopted from (Maier & Dandy, 2000) as cited in (Guerrero, 2016)**

ANN terminology	Statistical terminology
Input	Independent variables
Output	Predicted value
Training values	Dependent variables
Errors	Residuals
Training or learning	Estimation
Error function or cost function	Estimation criterion
Pattern or training pairs	Observations
Weights	Parametric estimates
Generalization	Interpolation and extrapolation

---

Waziri, et al., (2017) reviewed research papers to explore the applications of different ANN algorithms in the construction management domain. Their review indicated that ANN was found applicable in: cost prediction; optimization and scheduling; risk assessment; claims and dispute resolution; and decision making (Waziri, et al., 2017).

Bhokha and Ogunlana, (1999) using data of 136 high rise buildings from Thailand developed ANN model with 11 input node and 6 hidden layers.

Waziri, (2012) applied ANN with back propagation algorithm to develop a model for estimating project duration using data of 30 completed building construction projects from North Eastern part of Nigeria, where 70% of the data were randomly selected to be used as training dataset while 20% of the data were used as the testing set in which the performance of the ANN model was tested, while the remaining 10% were used as holdout sample (Waziri, 2012).

In Egypt Gab-Allah, et al., (2015) using data of 118 building projects, of which 105 used as training data set and 13 for testing developed ANN model for project duration prediction. Further developed a software that assist in predicting project duration (Gab-Allah, et al., 2015).

The high prediction efficiency of ANN models had been witnessed in literature where the coefficient of determination  $R^2$  value greater than 90 percent had been witnessed. ANN models from (Petruseva, et al., 2012) and (Golizadeh, et al., 2015) scored  $R^2$  value greater than 97 percent. The greater learning efficiency and ability to model the nonlinear nature of project durations make ANN models far more superior in explaining the variations in data sets. The major short coming of ANN models is that they lack the ability to show the relationship among the dependent and the predictor variables (Guerrero, 2016).

### **2.5.2.3 Simulation Based Models**

The technological advancement in the world of today made the application of simulation models for construction industry easier than ever. Simulation evolved and still achieving progress together with the development of computers (Bokor, et al., 2019). AbouRizk, et al., (2011) described construction simulation as a fast-growing field and defined it as a science of developing and experimenting with computer-based representations of construction systems to understand the interactions among components that made the system (AbouRizk, et al., 2011).

The wide application areas of simulation in the dynamic construction process span from simulating a single activity to modeling the entire construction interaction process. The main advantage of simulation model is creating the ability to view the uncertain aspects of a process and enhance a decision making ability of the management team involved in the

project. One of the areas where uncertainty handicapped its accuracy is the planning phase of a project (Bennett & Ormerod, 1984). Simulation technique is better applied when problems are characterized by uncertainty, when problems are technically or methodically complex, when flexibility in modelling logic and knowledge is required to formulate a model and when an integrated solution is required (Moradi, et al., 2015).

The complex interactions that exist between the different construction duration influencing factors made researchers to adopt simulation modeling for planning and scheduling a project. Bennett and Ormerod, (1984) developed a computer simulation program that was used to estimate the project duration at the tender stage using four different building projects, where it was recognized that the initial project duration estimate was too optimistic, while simulation based estimate was a closer approximation (Bennett & Ormerod, 1984). Dawood, (1998) attempted to develop a model that accurately model activity dependence and realistically predict project duration.

Similar applications of simulation to mimic and predict the project duration had been reported by; (Laptali, et al., 1995 ; Nguyen, et al., 2013 ; Bokor, et al., 2019). simulation models are either for a single trend of work e.g. masonry work by (Bokor, et al., 2019) and for entire building construction duration e.g. a computer program Project Duration Forecast (PRODUF) by (Ahuja & Nandakumar, 1985 ).

Regardless of the powerful capacity of simulation to capture the dynamic nature of the construction process there are barriers related with its development and application in the industry. AbouRizk, et al., (2011) indicated that simulation technologies application is mainly limited in the academic community for research purpose and the adoption of the method by the industry is in its infancy. Further three barriers are listed for lower implementation simulation in construction industry: the need for more sophisticated knowledge acquisition apart from the subject area in the field of software tools and applications; the more complex non formal development process that require a large bundle of numerical data; and the “black box” effect of simulation model where the evaluation of the final model is based on its outcomes (AbouRizk, 2010).

In this section it has been seen that simulation modeling approach is suitable for resembling the uncertain nature of construction duration and enable decision making capacity. The more complexity involved in the development process requiring a

multidimensional knowledge base together with the time taking evaluation and validation requirement made simulation method less applicable in the construction industry.

## 2.6 Modeling Construction Duration in Ethiopia

Research on modeling construction duration in Ethiopia given little attention. To the best knowledge of the researcher only five works were conducted in the past. Among those, one was done by a government firm and the other by students as part of undergraduate and graduate thesis work. As cited by Assefa, (2008) the pioneering attempt for using the BTC model concept were by Building and Transport Construction Design Authority (BaTCoDA). In 1987, BaTCoDA formulated completion time for building projects based on cost and the formula is given in equation 16.

$$T = 7to12C^{1/3} \quad (16)$$

Where;  $T$ - actual completion time in months,  $C$ - actual project cost in Ethiopian birr in millions unit, while 7 is for small sized buildings and 12 is for more complex building.

(Andinet, et al., 2006) as cited in (Assefa, 2008) using data of 29 public educational building projects in Ethiopia validated Brownmilow's time-cost relationship for the sample education sector construction projects as indicated in equation 17.

$$T = 7.06C^{0.47} \quad (17)$$

Where;  $T$ - actual completion time in months,  $C$ - actual project cost in Ethiopian birr in millions unit.

Apart from building projects the relationship between construction time and coast was found valid for road projects by (Assefa, 2008) and by (Birhane, 2015) for railway projects. Assefa, (2008) has validated the Brownmilow's Model by adopting it for estimating the construction duration of Ethiopian Federal Road Construction Projects. A research sample size of 33 federal road projects administered by Ethiopia Road Authority (ERA) under the categorization of International Contractor (IC) and Domestic Contractor (DC) based on the contractors who have undertaken the projects and on the adjusted and unadjusted cost basis was used. Project data about contract and final cost both in birr units, contract and

actual duration both in calendar days and length of the road in kilometer (km) were used for the analysis and construction time prediction models were developed.

Birhane, (2015) using project data of Addis Ababa Light Rail transit, Sebeta Mieso, and Mieso Dewanle railway projects attempted to develop a model to project duration based on actual executed amount and to check the validity of Bromilow's time-cost relationship model (Birhane, 2015). For all the three railway projects the BTC model was found valid and applicable. Further, Birhane, (2015) developed twelve regression equations using different mathematical transformations for enhanced performance of the predictive model and cubic transformation was found good in this regard.

The works by BaTCoDA and Andinet, et al., (2006) considered only project cost as a main influencing factor of building construction duration, which is a major gap of the previous studies. A recent graduate thesis work from Mekelle University by (Kifle, 2015) attempted to fill this gap by considering the effects of other project scope factors like; Gross floor area, Number of story, Height of building, and Average floor area.

Kifle, (2015) used a total sample of 56 (31 educational and 25 office) public building projects to validate the applicability of BTC model and develop a time prediction model. He used the mean absolute percentage error (MAPE) of the prediction model for identifying the performance of the models developed. Two qualitative variables were used in multiple linear regression analysis: project type (classes of Educational and Office building project) and Cost Level (Low Cost, Medium Cost, and High Cost). A case study research design that involved three government offices: Addis Ababa Science and Technology university project control office, Addis Ababa City Government Construction Agency and Federal Gov't Offices Construction Project Office was conducted to gather project information.

Bromilow time-cost principle was found to be moderately applicable with a coefficient of determination  $R^2$  value of 0.376 for the sample data set. An improved  $R^2$  value 0.487 for the time-cost relationship was achieved using the cubic regression model (CUB). The study also revealed a poor relationship between actual time and gross floor area for the case projects (Kifle, 2015). A good fit time model with  $R^2$  value of 0.513 using multiple linear regression analysis consist of three statistically significant variables; average floor area (Av. A), height of building (H) and cost level (CL) is given in equation 18 below.

$$Time = 599.003 + 200.464CL - 0.134AV.A + 7.838H \quad (18)$$

Where; *TIME* is Actual Time (Calendar days); *CL* is Cost Level factor [0 if low cost; 0.5 if medium cost; 1 if high cost]; *Av. A* is the Average Floor Area (m<sup>2</sup>); *H* – Height of building (m).

The minimum number of studies in modeling construction duration using statistical regression technique might be due to the unavailability of project information of previously constructed projects. The requirement of a large number of data set for generalizing results might be headache for researchers. Apart from willingness issue of respondents in providing information poor documentation of project history is evident in most contractors and consultants in Ethiopia.

## 2.7 Modeling Structural Works Construction Duration

Construction of a building can be regarded as the accomplishment of four major categories of work: Site preparation, excavation, structural framework, and finishing work. The site preparation concerned with: clearing the site; creating access to site; making temporary facilities; and site layout of the main building. Excavation is the work package associated with the removal of the natural earth to the desired level for setting the building position as per the design. The structural framework construction involves the erection of the building main structural skeleton the beam, column, and slab starting from the foundation of the building. The final work package is the finishing work which consists of: the internal and external wall system; the dressing and painting of the wall system; floor finishing; and electrical, mechanical, and plumbing works. The successful accomplishment of the four phases marks the end of the construction stage, therefore a careful planning and coordination among the four phases leads to a timely delivery of the building for its' intended purpose (Lee, et al., 2009).

The importance of the structural framework in a building construction is understandable. The importance came from the very fact that all the building load for which the building is designed to resist for is carried by it. Both the design and construction of the structural framework require care any negligence or mistake in worst case scenario might lead to the collapse of the entire building. The construction of the structural framework alone occupies approximately 50 percent of the entire construction duration of the building (Bang, et al., 2011). The increase in population together with scarce land resource made countries to

shift to high-rise building projects than the luxury villa housing typologies. The Upper structure work of this high-rise buildings is an element of activity that has the greatest weight with a percentage value of 21 percent of the total work weight (Husin, 2019).

Despite the high importance of structural framework construction phase of a building project there exist lack of a separate treatment of this phase from researchers specifically in modeling the duration aspect. It is clear that the total construction duration is the cumulative sum of every activity's duration in the critical path of a schedule, structural framework construction activities almost always present in the critical path. Apart from critical path activities that determine construction duration, the factors that lead to the realization of final duration might had been highly affected by a delay in certain separate phase of the construction. For example, a delay in the finishing phase due to lack of foreign currency might push the actual duration further from what was anticipated to be, such a scenario is better understood by separating the construction duration into separate segments.

Nkado, (1992) was the first to make a major premise that the uncertainty in building construction duration will be minimized, if an approximate estimate of durations of the various phases of a building is made, which is a benefit for all involved in the construction. He categorized the activities for the construction of new buildings into seven work groups as: setup, substructure, superstructure, cladding, mechanical and electrical (M&E) services, finishes and external works. Leaving the two work groups; setup and external works the five work groups were used to develop a construction duration analysis program of new buildings.

The structural work group can be separated into: Substructure work, which comprising all activities necessary to complete the ground works up to and including the ground floor slab, foundations, under-slab drainage, basement, etc. and the Superstructure work, include all activities necessary to erect the load bearing frame, including structural roof members (Nkado, 1992). Nkado, (1992) used a sample of 29 commercial, privately funded buildings data to fit a multiple linear regression model for the two structural work sub groups, equations 19 and 20 show the two models.

$$\begin{aligned} SUBST = & -9.295 \times BACCESS + 0.000561 \times VOLEXCAV \quad (19) \\ & + 0.00208 \times AREAGR D + 29.987 \times LONDON \\ & + 8.852 \end{aligned}$$

$$SUPERST = 0.491 \times HEIGHT - 11.056 \times CONSTN + 0.000663 \times AREATOTL + 14.827 \quad (20)$$

Where: AREAGR - is the ground plan area (m<sup>2</sup>); AREATOTL - is the total gross floor area(m<sup>2</sup>); BACCESS- is indicator of poor site access (poor, not poor); CONSTN- is indicator of type of structural frame (steel, concrete, other); HEIGHT- height from ground to eaves levels (m); LONDON- indicator of location (London, other); SUBST- substructure construction duration (calendar week); SUPERST- superstructure construction duration (calendar week); VOLEXCAV- volume of excavation (m<sup>3</sup>).

In Hong Kong Chan & Kumaraswamy, (1999) used the same approach as Nkado, (1992) to model the construction duration of 56 standard 'Harmony' type apartment blocks. The primary work packages in the building process determined were namely, piling, pile caps/raft, superstructure, E&M services, finishes and their respective sequential start-start lag times. Unlike Nkado, (1992) the division of the structural works only considers the superstructure group, the substructure as the blocks are standard all use pile foundation and it found appropriate to model the pile driving duration (Chan & Kumaraswamy, 1999b). Equation 21 shows the developed multiple regression model for superstructure construction.

$$\begin{aligned} SUPERSTR &= 4.0511 + 0.1316 HEIGHT \\ &+ FACADE (0 \text{ for with facades; } 1.6216 \text{ for without facades}) \\ &+ LABOUR (-0.8319 \text{ for readily available; } 0 \text{ for somewhat difficult}) \end{aligned} \quad (21)$$

Where:FAÇADE- indicator of presence of precast facades; HEIGHT- height (m); LABOUR- labour availability; SUPERSTR- duration of superstructure (months).

Another approach to modeling structural works duration is suggested by Yun and Kim, (2012), where they identified Structural work to consist of formwork, rebar work, concrete work, mechanical work, and electronic/communication work. By modeling the typical duration of these activities, they were able to estimate the structural framework construction duration of a case study building in Korea (Yun & Kim, 2012). Similar approach was used in Korea by Lee, et al., (2009) and in Vietnam Nguyen & Nguyen, (2013 ). Table 2-7

below summarizes the key characteristics of the models conducted in relation with structural work.

**Table 2-7 Comparison of previous structural work duration estimation models**

Distinguishing features	(Nkado, 1992)	(Chan & Kumaraswamy, 1999)	(Lee, et al., 2009)	(Bang, et al., 2001)	(Yun & Kim, 2012)	(Nguyen & Nguyen, 2013 )
Country	UK	Hong Kong	Korea	Korea	Korea	Vietnam
Purpose	Modeling total project duration	Modeling total project duration	Uncertainty in duration	time shortening	Optimum construction method	Relation b/n labour productivity & floor area
Method used	Questionnaire survey	Questionnaire survey	simulation	Case study	Case study	Case study
Structural work studied	Substructure & superstructure	Pile foundation & superstructure	Floor area	Typical floor	Superstructure	Typical floor

It had been seen that previous studies attempt to model various influencing factors of structural framework construction from different perspective and study purpose. It is evident that structural framework construction is an outdoor activity which is highly dependent on weather condition (Lee, et al., 2009). The learning curve of the work crew is also another major influencing factor which determines the duration (Nguyen & Nguyen, 2013 ). The technical aspect of structural work duration mainly depends on the concrete activities, which determines the formwork remaining period. The strength gaining and curing duration required involved in setting up the critical duration part of structural works (Bang, et al., 2011; Yun & Kim, 2012).

The observation of Mensah, et al., (2017) in a comprehensive literature review of previous studies in modeling construction duration both in building and other civil structures indicated that there is no inclusion of Bill of Quantity (BOQ) items as influencing factors. Here also in modeling studies of structural framework, except for Nkado, (1992) who used the volume of excavation as an influencing factor in duration model the other models don't include BOQ items in model. Structural frame work construction BOQ elements include but not limited to quantities of: formwork, reinforcing bar and concrete had not been considered in previous studies. This research attempted to fill the gap in this area by including the three previously mentioned BOQ items together with quantity of excavation.

It is clear from the above review that the structural framework duration can be studied from two perspectives: either by dividing it into two separate work packages namely,

substructure and superstructure e.g. (Chan & Kumaraswamy, 1999b; Nkado, 1992), or studying the individual activities involved namely, formwork installation, rebar placing, and concreting either in a typical floor or a unit area specified in a building e.g. (Lee, et al., 2009; Yun & Kim, 2012). For the current study the division of structural works into the two work packages (substructure & superstructure) is found to be convenient.

## 2.8 Construction Duration Model Prediction Performance Measures

Determination of the accuracy of predictive models is critical because predictive models have been increasingly used across various disciplines and predictive accuracy determines the quality of resultant predictions (Li, 2017). Li, (2017) indicated that Pearson product-moment correlation coefficient (R) and the coefficient of determination ( $R^2$ ) even though commented to be biased, insufficient and misleading they are among the most widely used measures for assessing predictive models for numerical data. Neither R nor  $R^2$  are measures of predictive accuracy rather they are measures of the goodness-of-fit between y and x.

The difference between the observed values and predicted values produce the required predictive accuracy of a model. Most of the study referred except few used the observation data samples that were used to develop a model to measure the predictive accuracy Li, (2017), however, indicated that this accuracy measure how well the model fits the training samples, thus it is not measuring the predictive accuracy. Predictive accuracy based on the differences between the predicted values for, and the observed values of, new samples (e.g., validation samples) should be used instead (Li, 2017).

Four simulated scenarios conducted by Li, (2017) showed that both R and  $R^2$  failed to measure the accuracy of a model and provided a misleading outcome, therefore Li, (2007) proposed two options to R and  $R^2$  these are: Variance explained by predictive models based on cross-validation (VEcv) and Legates and McCabe's efficiency ( $E_1$ ) and indicated that they are free of the limitations associated with the more frequently utilized measures R and  $R^2$ . Further from the simulated scenario graphs he concluded that both VEcv and  $E_1$  are equally interpretable. The key differences between them are that 1) VEcv explains the percentage of the variance of validation samples, while  $E_1$  explains the percentage of the sum of the absolute differences; and 2) VEcv is quadratically related to the differences between the predictions for and the observations of validation samples, while  $E_1$  is linearly

related to the differences. These two measures were also found to be independent of the sample size and data variation.

Equation 22 and 23 provide  $VE_{cv}$  and  $E_1$  formulas respectively;  $A_i$  is the actual observed value,  $\bar{A}$  the mean of the observed values,  $P_i$  is the predicted value from the fitted model, while  $n$  is the number of samples in the validation data set.

$$VE_{cv} = \left( 1 - \frac{\sum_1^n (A_i - P_i)^2}{\sum_1^n (A_i - \bar{A})^2} \right) \quad (22)$$

$$E_1 = \left( 1 - \frac{\sum_1^n |A_i - P_i|}{\sum_1^n |A_i - \bar{A}|} \right) \quad (23)$$

Performance evaluation of models considered as an interdisciplinary research problem. Performance metrics (error measures) are vital components of the evaluation frameworks in various fields, they are defined as; *“a logical and mathematical construct designed to measure how close are the actual results from what has been expected or predicted.”* (Botchkarev, 2019).

In the literature six different types of performance evaluation metrics used either alone or in combination the other ones the highest combination being the one by (Love, et al., 2005) where three of them used to measure the accuracy the construction duration forecasting models. The list of performance metrics observed in literature provided below:

- Absolute percentage error (APE),
- Mean absolute error (MAE),
- Mean absolute percentage error (MAPE),
- Mean square prediction errors (MSPR)
- Percentage error (PE),
- Root-mean-square error (RMSE),

As observed in the literature almost all had not put a justification for the choice of a particular performance metric for use. Even though there are other performance metrics the discussion is limited to those used in previous studies, with the main aim of identification of the weak and strong parts of performance metrics so that a more reliable metric is selected for the current study.

Botchkarev, (2019) indicated that some metrics are more popular than the others, the three top most common metrics that came up in many studies that were performed over a timeline of 25 years are: mean square error (MSE) (or root MSE (RMSE)), mean absolute error (MAE) and mean absolute percentage error (MAPE). A shift in the preferences towards a given metrics observed in the study. Particularly in the 1980s the preferred metrics was, MSE/RMSE. While this preference moved toward MAPE in the 1990s and 2000s (Botchkarev, 2019).

In several different disciplines researchers strive to find a best metric and develop new metrics either for a task specific use or by combining existing metrics, but still, no consensus on the “best” metric has been achieved (Botchkarev, 2019).

The point distance between actual ( $A_j$ ) and predicted value ( $P_j$ ) can be found by the application of mathematical operations as; error (magnitude of error) =  $A_j - P_j$ , absolute error =  $|A_j - P_j|$ , or squared error  $(A_j - P_j)^2$  (Botchkarev, 2019). Each point distance estimations have their pros and cons in the final reporting of performance (Botchkarev, 2019). From the identified six performance metrics above three of them use absolute error form for distance calculation (APE, MAE, MAPE), two of them squared error (MSPR, RMSE), and one magnitude error (PE). Magnitude error a simplest form of calculating distance might lead to mutual cancelation of opposite signed errors. Both absolute and squared errors remove this mutual cancelation of opposite signed errors. Squared error penalizes extreme errors and outliers and minimize the effect of the smaller ones (Botchkarev, 2019).

Different authors and researchers identified the pros and cons of the various performance metrics used in different disciplines e.g. (Steyerberg, et al., 2009; Myttenaere, et al., 2015; Guerrero, 2016; Tian, et al., 2016 ). Li, (2017), however indicated that the most performance metrics cannot tell how accurate the models are, rather, they are error measures and not accuracy measures, so they can only tell which model produce less error

but they are unable to tell how accurate the models are. VECv is an accuracy measure that is unit/scale, data mean and variance independent. It is an accuracy measure of predictive models and thus their predictions, and it provides a universal tool to assess and directly compare the accuracy of predictive models for any numerical data of various unit/scale, mean and variation from any disciplines.

The review indicated that statistical measures used as an accuracy measures in previous studies are really not indicating what was supposed to be indicated, detailed recent works of Li, (2017) and Botchkarev, (2019) paved a new road for evaluation of the predictive models of the current work. Therefore statistical measures R and  $R^2$  will be used to measure the goodness-of-fit between dependent variable (y) in this paper the construction duration of structural frame work and independent variable (x); selected project scope and bill of quantity items of structural framework. While, the root mean squared error (RMSE) used to measure accuracy and selection of model as suggested by Guerrero, (2016).

## **2.9 Summary of Literature Review and Research Gap Identified**

### **2.9.1 Summary of literature review**

The literature review conducted in this chapter of the thesis helps to analyze the findings and suggestions from previous related studies. Further it was one of the main tool used in discovering research variables together with subject matter knowledge in the research specific area. As indicated in the introduction of the chapter it was attempted to be specific to include researches conducted on building projects only. This helped to identify project specific findings in our topic: construction duration.

A separation of the general topic into three main sections; 1) building construction duration influencing factors, 2) modeling building construction duration, and 3) modeling building structural framework construction duration, was used to discuss previous related works. Further a supporting section focusing on predictive performance measures of duration forecasting models were done to analyze and suggest such a measure for the current work.

Construction duration influencing factors were discussed by separating them into two broad classes; project related and non-project related factors. Project related factors defined as factors that are born either from the physical nature of the building or the nature of the stakeholders and the organization of project team. While non-project related factors

are those which are outside the project and the project team have little or none control over such factors, and also are characterized by high uncertainty. The identification of construction duration influencing factors had been indicated as an important step for devising a predictive model for future estimation use. Further also identification of such factors enables to identify their magnitude of influence and construct a method to minimize their influence. Different researchers used different grouping of individual factors and attempted to rank such factors. But it was evident from different previous studies there is no common categorization to apply for and an agreed rank of such factors. Therefore, it is concluded to include project scope factors specially those which are quantitative (for which a meaningful numerical value can be assigned) into this study of structural framework construction duration.

Four modeling approaches; Linear regression, Fuzzy logic, ANN, and Simulation from previous studies identified and discussed into two separate sections; linear regression in a separate and wider consideration and the other three in one category. Much emphasis was given to mathematical models which are based on linear regression since the current paper also used this approach to model the construction duration of structural framework of buildings.

The first mathematical relationship used for modeling construction time used a single predictor variable which was construction cost where it was assumed to be a measure of the project size. Later studies identified the weakness of such a relationship and especially the use of cost as a predictor variable during earlier stages of planning where cost is not known. Therefore, number of variations to the original model were used in previous studies and discussed here. Specially the use of multiple linear regression for inclusion of more than one influencing factor both qualitative and quantitative produce models with a capability of describing a larger portion of variation in construction duration.

Predictive models' ability to describe future observations with less error is found to be a main measure of performance for such models. The use of six error measures in previous studies was discussed. Correlation coefficient, coefficient of determination and Root mean squared error (RMSE) are identified to check model accuracy and during selection.

### **2.9.2 Research gap identified**

A number of researchers, but few analyzed and modeled construction duration by segmenting the entire phase into separate phases. The structural framework construction phase which is considered as a major portion in building construction gained little attention from researchers. Existing literatures used two approaches to model the duration; either subdividing the structural framework phase into substructure and superstructure works or modeling individual activities like formwork, rebar work, and concreting work. This research used the first approach of classifying into substructure and superstructure. Clear definitions of terms used in this regard are also provided together with a diagrammatic illustration of the framework division.

The gap identified in previous construction duration models by Mensah, et al., (2017) was found true for structural framework duration models too, and is planned to include four BOQ items in this study. Further with the aid of previous studies and engineering judgement the following variables are hypothesized as structural framework influencing factors: gross floor area, height, number of floors, quantity of concrete, quantity of rebar, and quantity of formwork. A full list of measured variables provided in chapter three and four of this thesis.

## CHAPTER 3 RESEARCH METHODOLOGY

### 3.1 Introduction

The research methodology section explains the materials, techniques, samples, data, approaches, theoretical frameworks used in the study. How the research was carried out together with a clear presentation of the procedure followed in the preparation and analysis of the data is also included.

### 3.2 Research Design

Kumar, (2011) defined research as a process of collecting, analyzing and interpreting information to answer the questions of interest. The procedure of identification of the clear path toward answering the research problem is termed as research design. A research design as defined by Kumar, (2011), is a plan, structure and strategy of investigation. The purpose should be convincing both to the reader and the researcher that the path is correct and will yield a valid answer to the research problem.

Here in this subsection, discussions on; type of research chosen for the problems and the hypotheses set during the first chapter are presented. As part of the research design, detailed information about; study population, study sample, data collection and determination of the appropriate size of respondents together with other relevant issues are discussed separately in the following sections. Such separate discussion was assumed to provide a better understanding of the methodology applied and helpful for future replication of the methodology in other similar studies.

#### 3.2.1 Type of research

Research came in different types according to different classifications, and it is necessary for a researcher to choose among those where his/her research tend to fit in. Three classification systems from Kumar, (2011) used to identify the appropriate categories for this research.

The first classification is according to the application of findings of a research two types of research exist; pure and applied research. Kumar, (2011) defined pure research as the process of developing and testing challenging theories that may or may not have practical

application. While in applied research the research techniques, procedures and methods that make up the research methodology are applied to the collection of information. This research is an applied research as the results of the study, which are predictive models are going to be used by real estate developers for predicting the construction duration of structural framework of building projects.

The second classification system categorizes research works based on their objective as; descriptive, explanatory, exploratory, and correlational. This research fits under correlational sections where, the objectives clearly indicated the investigation of the kind and degree of relationship among structural works duration and identified influencing factors: project scope related factors, quantity of works, and chosen construction methods.

The last and third classification system categorize research works according to the enquire mode, where quantitative and qualitative methods exist. This research used quantitative enquire mode to achieve the objectives and answer the proposed research hypotheses.

Further, the research is distinguished as a non-experimental survey research based on the data collection method used to gather information about the study sample and identified variables. Therefore, it can be concluded that this research is an applied, non-experimental, quantitative, correlational, survey study.

### **3.3 Population of the Study**

Population is a theoretically identified aggregation of elements of study (e.g., person, object) (Dattalo, 2008). In this research the population of the study are made of the building projects undertaken by private real estate developers based in Addis Ababa. With this regard, the exact size of the population was not identified due to difficulty to get information about the total number of such projects.

### **3.4 The Study Sample and Sample Size Determination**

#### **3.4.1 Study sample**

The sample of a study is the subset of the population. The study population comprises of all building projects constructed by real estate developers based in Addis Ababa. The population was only limited on geographical area. Therefore, building projects of three

different categories; apartment and villa buildings constructed for residence purpose and commercial buildings for business offices are taken from the study population through a sampling procedure of respondents (real estate developers) to make the sample of the study.

### **3.4.2 Sample size determination**

Dattalo, (2008) indicated sample size determination as an important and difficult step in planning an empirical study. Previous studies that used regression modeling made little concern about the sample size and further with exceptional few none provide a justification for the appropriateness of the sample size for their study.

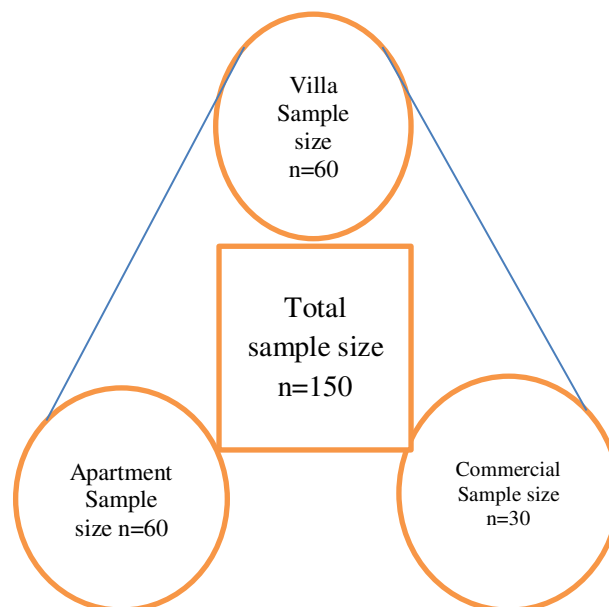
Kumar, (2011) indicated that the question of “*how big a sample is enough?*” is always a challenge for a researcher. He suggested that the answer depend on the answer to two other questions; what you want to do with the findings of your research? and what type of relationship do you want to establish? But he concludes the task of deciding sample size as the “*the larger the sample size the more accurate your estimates*” (Kumar, 2011 ). Edmonds and Kennedy, (2017) pointed that increased sample size guarantee a decreased sampling error, further the larger the sample size the more likely the results to represent parameters of the population of the study. Statistically significant result is only achieved with large enough sample sizes (Brooks & Barcikowski, 2012 ). The question of “how large is enough?” is also raised in this research.

The major deference arising between the use of parametric and non-parametric tests is that parametric testes assumes the population distribution to be normal while their non-parametric tests are distribution free. Mendenhall, et al., (1993) as cited in Guerrero, (2016) indicated that a sample size of 30 subjects is sufficiently large for a normal distribution approximation (Mendenhall, et al., 1993). A same number of sample size is suggested by Etikan, et al., (2016) for use if the research has a survey design. Studies with small sample size less than 30 participants provide imprecise and unreliable estimates (Barton & Peat, 2014).

Brooks and Barcikowski, (2012) identified three approaches for determination of sample sizes of studies that use multiple regression: conventional or rule of thumb method, statistical power approaches, and cross-validation or data splitting method. Conventional methods provide a minimum sample size together with an adjustment to the number of

predictor variables. Different authors suggested different minimum sample sizes, which will make a researcher indifferent between any set of alternative. One researcher suggests 15 subjects and another 50 as sufficient justifying such an alternative to become difficult. The second method uses sample size suggestion based on the probability of rejecting a null hypothesis, when the null hypothesis is indeed false (Brooks & Barcikowski, 2012 ). The main weakness of sample size based on statistical power is its' inability to ensure generalization of a regression equation to other samples. The third method requires a large set of sample for two purposes for model development and model validation purpose and it is refereed as either data splitting or cross-validation technique (Snee, 1977; Claudy, 1978 ; Brooks & Barcikowski, 2012 ; Sarstedt & Mooi, 2014).

For this research, a minimum sample size of 30, a requirement of parametric test utilized and in order to have sufficient sample size for both model development and validation purpose, a sample size of 150 was assumed good enough with a minimum size of 30 for each project type. below shows the composition of the sample. Figure 3-1 shows the study sample composition used for determination of the representative respondents.



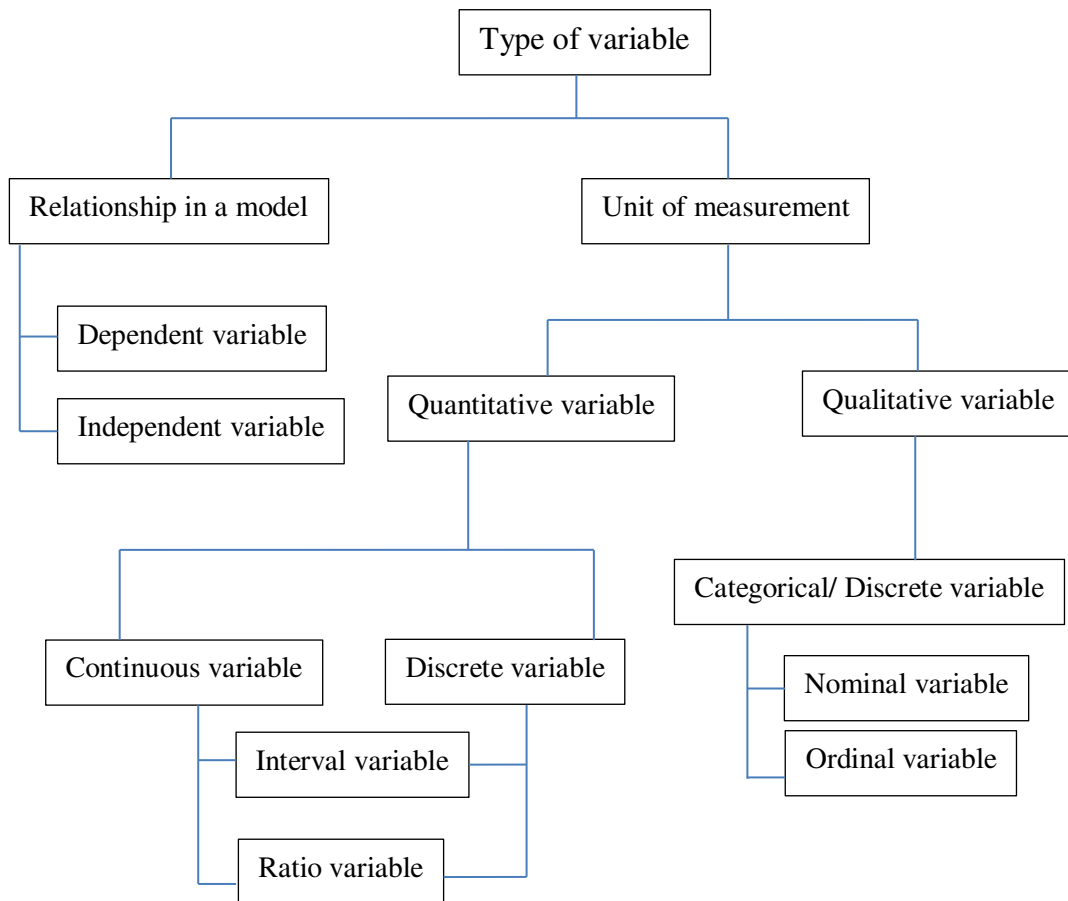
**Figure 3-1 Study sample composition and sample size**

### 3.5 Identified Variables

A concept that can be measured and for which different values are attached is termed as a variable (Kumar, 2011 ). In this research, during the literature review concepts related with

construction duration of building projects in general and specifically structural framework duration were identified. Concepts need to be converted into variables either directly or using indicator variables that can differentiate between different categories.

Variables can be of different type depending on their use in a given study. Here for the purpose of selecting appropriate statistical tests, there is a need to differentiate between different types of variables. A classification of a variable system applicable for this study is shown in Figure 3-2 below.



**Figure 3-2 Classification of variables used for the study**

The list of all identified dependent and independent variables tabulated in Table 3-1 and Table 3-2 respectively below show their respective symbols for use during data analysis.

**Table 3-1 Identified dependent variables of the study**

No.	Identified dependent variable	Unit	Symbol	Type of variable	
				measurement scale	Unit of measurement
1	Total structural works construction duration	Days	<i>D_Total</i>	Ratio	continuous
2	Superstructure works construction duration	Days	<i>D_Super</i>	Ratio	continuous
3	Substructure works construction duration	Days	<i>D_Sub</i>	Ratio	continuous

**Table 3-2 Identified independent variables of the study**

No.	Identified independent variable	Unit	Symbol	Type of variable	
				Unit of measurement	Measurement scale
1	Total number of floors	Number	<i>T_Floors</i>	Discrete	Ratio
2	Number of floors below ground	Number	<i>B_Floors</i>	Discrete	Ratio
3	Number of floors above ground	Number	<i>A_Floors</i>	Discrete	Ratio
4	Total Building height	m	<i>T_Height</i>	Continuous	Ratio
5	Typical floor height	m	<i>Ty_height</i>	Continuous	Ratio
6	Total gross floor area	m <sup>2</sup>	<i>T_GFA</i>	Continuous	Ratio
7	Typical floor area	m <sup>2</sup>	<i>Ty_area</i>	Continuous	Ratio
8	Quantity of excavation	m <sup>3</sup>	<i>Q_excv</i>	Continuous	Ratio
9	Quantity of formwork	m <sup>2</sup>	<i>Q_form</i>	Continuous	Ratio
10	Quantity of reinforcing bar	Kg	<i>Q_bar</i>	Continuous	Ratio
11	Quantity of concrete	m <sup>3</sup>	<i>Q_conc</i>	Continuous	Ratio
12	Type of foundation	none	<i>Found_ty</i>	Categorical	Nominal
13	Type of slab	none	<i>Slab_ty</i>	Categorical	Nominal
14	Concrete mixing method	none	<i>Conc_method</i>	Categorical	Nominal
15	Use of tower crane for construction material transportation	none	<i>Tower_C</i>	Categorical	Nominal
16	Shoring & scaffolding system used for formwork	none	<i>Form_sys</i>	Categorical	Nominal
17	Use of concrete pump	none	<i>Conc_pump</i>	Categorical	Nominal

The identification and categorization of variables together with their appropriate measurement scale has a vital use during the analysis since the statistical software dependent on the user defined properties of variables.

The variables were identified based on inputs from previous study and subjective judgement of the researcher. The influencing or independent variables were categorized as project scope related factors, quantity of work items, and construction method. The concept of project scope related factors as applied for building project were discussed in chapter 2 in more detail. Variables of quantity of work item are new additions to the concept of duration influencing factors which were first identified as a gap in modeling duration studies by (Mensah, et al., 2017). While the construction methods variables are also based on the knowledge of the subject area of structural frameworks. Table 3-3 Below shows a categorization of the seventeen identified independent variables.

**Table 3-3 structural framework duration identified influencing factors categorization**

Category of structural framework influencing variables		
Project scope related	Quantity of work item	Construction method
<i>T_Floors</i>	<i>Q_excv</i>	<i>Found_ty</i>
<i>B_Floors</i>	<i>Q_form</i>	<i>Slab_ty</i>
<i>A_Floors</i>	<i>Q_bar</i>	<i>Conc_method</i>
<i>T_Height</i>	<i>Q_conc</i>	<i>Tower_C</i>
<i>Ty_height</i>		<i>Form_sys</i>
<i>T_GFA</i>		<i>Conc_pump</i>
<i>Ty_area</i>		

### 3.5.1 Operational definitions of identified variables

The study variables (three dependent and seventeen independent variables) might bear different meanings to different readers. It is found helpful to present operational definitions for clarity purpose and full understanding of the interpretations of the values of variables in the analysis chapter of this thesis.

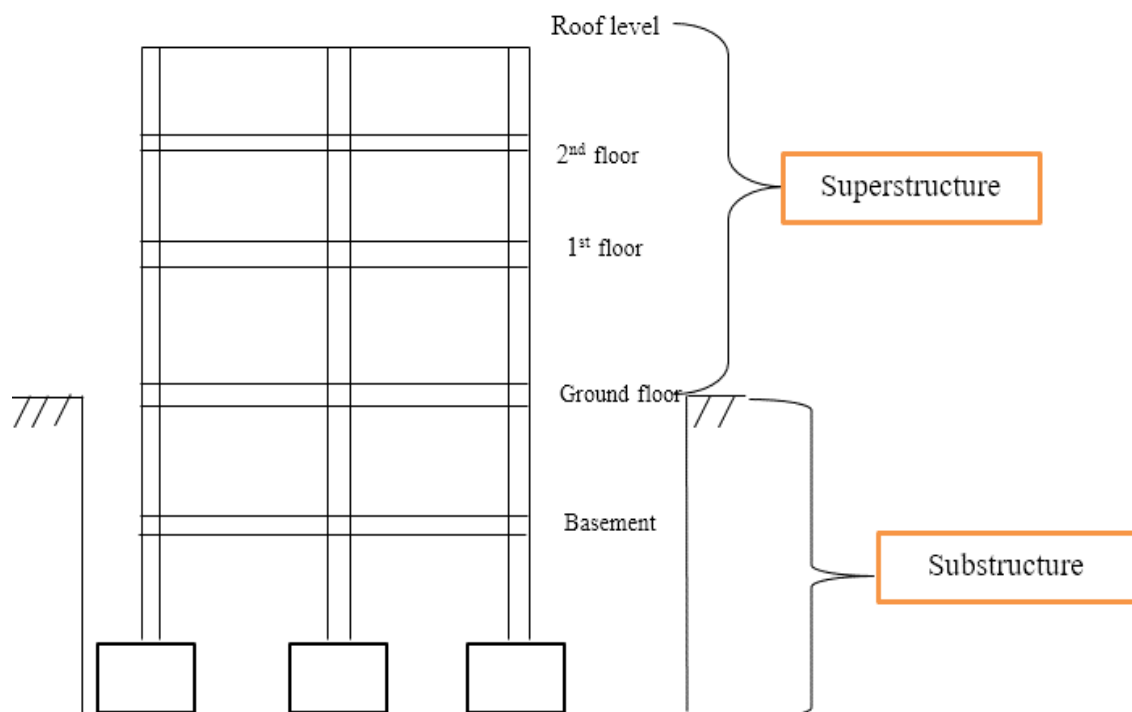
The structural works involved in building construction were first divided in to two categories of work as; substructure and superstructure works. This study recognized the following sub tasks and activities as part of the *structural works group*:

- All activities involved in erecting foundation of a building, here even though excavation is included among influencing quantity of work items variable doesn't mean it is part of the structural framework.

- All activities required for erecting the foundation column, beam, slab, and basement of the building up to the ground floor level.
- All activities part of the upper structure starting from the columns and/or load bearing walls rising from the ground floor up to the roof level.
- All activities for erecting the building circulation facilities i.e. stairs and elevator shaft.

The term *all activities* as used here includes, but not limited to: formwork preparation, installing in place & removal; rebar preparation & installing; concrete mixing, delivery, pouring, compacting & curing.

*The substructure work group:* includes all the activities from the foundation to the ground floor slab, while the activities from the columns and/or load bearing walls rising from the ground floor to the beam, column, slab system up to the roof level forms the *superstructure work group*. Figure 3-3 below shows the operational classification used in the study of structural works.



**Figure 3-3 Operational classification of structural work groups.**

Three different construction durations were identified as dependent variable for the study their operation definition are:

- *Substructure works duration ( $D_{Sub}$ )*- the duration spent for construction of structural elements from the foundation up to and including the ground floor slab.
- *Superstructure works duration( $D_{Super}$ )*- the duration elapsed for construction of structural elements starting from the ground floor slab column rising to the first floor slab up to the roof beam construction.
- *Over all structural works duration ( $D_{Total}$ )*- the duration took by the construction of the entire structural framework system from the foundation up to the roof beams.

The variables under construction method were developed from subject matter knowledge are qualitative variables but for the purpose of analyzing their influence appropriate categories for each was created. Table 3-4 below shows the categories created for the analysis purpose.

**Table 3-4 Qualitative variables used with their identified measurement categories**

Construction factor	method	Symbol	Identified categories for measurement
Type of foundation		<i>Found_ty</i>	Isolated/combined footing, Raft/Mat foundation, Pile foundation
Type of slab		<i>Slab_ty</i>	Conventional 1 way or 2 way, Flat slab, Ribbed slab
Concrete mixing method		<i>Conc_method</i>	On site mixing, Ready mix delivered with truck
Use of tower crane for construction material transportation		<i>Tower_C</i>	Yes, No
Shoring & scaffolding system used for formwork		<i>Form_sys</i>	Eucalyptus wood, Metal, Combination
Use of concrete pump		<i>Conc_pump</i>	Yes, No

The qualitative factors of construction method factors except for the two binary (yes, no) categorical variables; measuring the use of tower crane and concrete pump, the other factors categories are not limited to those used for measuring them. Different types of slab, foundation, mixing method, shoring and scaffolding do exist other than those proposed here, therefore readers should understand this and keep in mind.

Except for the variable gross floor area, the other variables in both project scope related and quantity of work are assumed to hold a self-descriptive meaning and are understandable by readers of this work. Therefore, an operational definition is provided only for this variable.

- *Total gross floor area*: is the sum of all areas on all floors without deduction for stairs, elevator and building services shafts, corridors, the thickness of interior walls, columns or other features

## 3.6 Data Collection

### 3.6.1 Instrument

After the identification of the response and influencing variables together with determination of appropriate sample size the next stage in research is the preparation of appropriate data collection instrument for the identified variables. A research instrument is anything that is used for the purpose of collecting information (Kumar, 2011). Kumar, (2011) divided information gathering approaches into two; use of available information and collection of required information. The selection between the two approaches depend on the type of research, availability of required data, reliability of data required.

The literature review carried out in chapter two showed the application of both approaches for duration modeling purpose. The use of available data was noticed only on two works: (Stoy, et al., 2007) used data from BKI data base in Germany while, (Guerrero, et al., 2014) in Spain used a database that contains more than 300 projects with different uses. The use of available data is both time and cost saving leaving the researcher with plenty of time for data analysis and discovering any type of hypothesized relationship. The limited availability of historical data lead to the adoption of questionnaire survey instrument by majority of the studies reviewed.

As part of the data collection instrument process a data sheet for gathering the values of the identified variables of the three sample categories; apartment, villa, and commercial building developed. A sample data sheet can be refereed in the Appendix section of the paper. Basic recommendations for preparation of questionnaire were utilized from different sources. To measure the effects of qualitative independent variables categories were provided as close ended question. While the rest of the questions for quantitative variables are open ended. Together with the data sheet a covering letter indicating the objectives of the survey and a one-page illustrative diagram of typical structural framework with definitions provided for clarity and to increase the assumed response rate. A total of 300 data sheets were prepared and identified with code for apartment and villa sample project and 180 data sheets for commercial building project sample. A survey package containing 15 pages (1-page cover letter, 1-page illustration, and data sheet of 5 pages for apartment, 5 pages for villa, and 3 pages of commercial data sheet.) was sent for each sampled respondent real estate developers. Respondents were coded with ID and

prepared a sheet that identify the data sheet codes together with the respondent ID. These ensure to check all collected data sheets systematically.

### 3.6.2 Respondents size determination

Questionnaire as a primary source of data embeds low response rate as a major weakness. Kumar, (2011) stated that one should found himself lucky for a response rate of 50%. Further he noted the following issues as causes for much lower response rates:

- People fail to return questionnaires;
- Interest of the respondent in a topic of the study;
- Layout and length of questionnaire;
- Quality of the letter in explaining the objectives of the study.

Involving a larger set of respondents might help to gather sufficient number of samples with a lower level of response rate, but the question of cost and time came to the picture with this approach. The required number of respondents was calculated by first assuming a lower response rate of 20% for apartment and villa project which requires distribution of five questionnaires for receiving one response. Even a much lower response rate of 17% for commercial buildings taking into consideration the lower rate of involvement in construction. Then some simple procedural calculations were carried out to came up with the required number of respondents of 60 for obtaining a sample of 150 projects. Table 3-5 below show the detailed steps followed.

**Table 3-5 Calculation of size of respondents (real estate developers)**

Building category	Apartment	Commercial	Villa
Required sample size: <i>A</i>	n=60	n=30	n=60
Assumed response rate: <i>B</i>	20%	17%	20%
Sample size required from a single respondent: <i>C</i>	5	3	5
Total questionnaire to be distributed: $D = (A*100)/B$	300	176.5 round up to 180	300
<b>Total number of respondent required: <math>E=D/C</math></b>	<b><u>60</u></b>		

### 3.6.3 Sampling Procedure of Respondents

Sampling procedure is the technique used to select elements of a study from identified set of population. The main objective in sampling procedure is obtaining a sample representative of the study population. The accuracy of the research findings depends upon the sample selection procedure used (Kumar, 2011 ).

There are two types of sampling; Random or Probability sampling and Non-random or Non-probability sampling. In Random/ Probability sampling all elements in the study population have equal and non-zero chance of being part of the sample, while this is not possible in Non-random/ Non-probability sampling techniques where each element has different chance of being part of the study. Random sampling is the most common and reliable means of achieving an unbiased sample from a population. Non-random sampling is used when the number of subjects in a population is unknown or cannot be individually identified (Kumar, 2011 ; Edmonds & Kennedy, 2017).

In this research real estate developers in Addis Ababa were sampled from which data about the sample; i.e., building projects of three categories; apartment, commercial and villa obtained. This research adopted both types of sampling techniques (Random and Non-random sampling) to select respondents that provide information about the sample. A total of 60 real estate developers as respondents were sampled, out of which 30 were sampled using a simple random system while the other half using purposive or judgmental type of non-random sampling technique. The 60 real estate developers were found sufficient after considering a response rate for the data collection instrument designed and it is discussed under respondents' size determination section previously, here now the procedure adopted in selecting the respondents is discussed.

The simple random sampling was conducted on Microsoft Excel software. First list of 124 real estate developers in Ethiopia was identified as a sampling frame from the Federal Democratic Republic of Ethiopia (EFDRE) National telephone directory yellow page under the real estate category (Ethiotelecom, 2016). This list was further reduced to 102; where 13 were found on non-probability sampling list and 9 were discarded since they are not in Addis Ababa. The list of these 102 real estate developers were written on excel sheet and numbered from 1 to 102 and the excel function *RANDBETWEEN (bottom, top)* used to return a random number between 1 and 102 and randomly drawn list of the first 30 real estates were taken as a random sample.

Purposive or judgmental sampling is a non-probability sampling technique, which involves the use of the researcher's knowledge of the population in terms of research goals (Edmonds & Kennedy, 2017; Sharma, 2017). Samples were selected based on the criteria of the reputation of real estate developers in the industry. This reputation was attempted to be confirmed by using internet search terms to retrieve list of "*best real estate*

*developers in Addis Ababa*” further also well acknowledged companies also made the list based on our judgment. Further a strong assumption made about their willingness to provide information about their previously conducted projects taken from their collaboration in previous other studies conducted in Addis Ababa University was also another criterion to select the other 30 respondents.

### **3.7 Statistical Data Analysis Approach**

In this subsection details of the procedure followed for the statistical analysis of data gathered about the identified variables through the questionnaire survey is presented. The analysis was done using Statistical Software Package for Social Study (SPSS) software version 23.

#### **3.7.1 Data preparation**

The starting step for data analysis was the preparation of data for analysis. The very basic moto in using software for data analysis is *“garbage-in garbage-out”* therefore, the following tasks were conducted:

- Preparation of excel work sheet for transcribing values of identified variables;
- Transcribing the data from returned questionnaires on Excel worksheet;
- Identifying non returned data sheets and reporting the response rate;
- Exporting data from excel to Statistical analysis software; i.e., to SPSS;
- Checking for the required assumptions of selected statistical analysis methods;

#### **3.7.2 Statistical methods used**

The analysis of data for testing hypothesis, answering research question and meeting the general objective of developing mathematical model for future prediction of structural framework construction duration required us to use statistical methods; correlation and linear regression analysis methods and non-parametric tests to compare groups of construction method.

##### **3.7.2.1 Correlation analysis**

The research questions required determination of relationship among the identified dependent and independent variables. Correlation coefficient describes how closely two variables are related, i.e., the amount of variation in one variable explained by another

variable. Pearson's correlation coefficient ( $r$ ) used for this purpose. The choice of correct statistical test depends on the researcher careful identification of variable type in addition to the research design (Edmonds & Kennedy, 2017). The research variables as presented earlier are either continuous or categorical variables. Pearson correlation is applicable only for two continuous variables (Barton & Peat, 2014).

Two basic assumptions of Pearson correlation require the two variables to be linearly related and have normal distribution. The linearity assumption was checked using scatter plots of the two variables. While the normal distribution assumption was checked using Kolmogorov-Smirnov statistical test and visual display of histograms of variables of interest. In cases where these assumptions failed, data transformation used and rechecked the assumptions. Data transformation is the use of mathematical operation to change original data to normal distribution.

### **3.7.2.2 Linear regression analysis**

Correlation coefficients provide information about existence of relationship but, fail to quantify it. Two types of linear regression analysis were used to build predictive models; simple linear regression (SLR) and multiple linear regression (MLR). SLR was used to identify the magnitude of relationship between two variables. It was used to identify the extent of variability explained by a single influencing factor on the dependent variable (structural works durations). MLR help to identify the extent to which more than one independent variables predict the dependent variable structural works duration. As the name suggest linear regression assume that, two variables are linearly related, therefore, existence of such relationship was examined using scatter plots and perform data transformation when found applicable to linearize the variable.

### **3.7.2.3 Non-parametric tests**

Non-parametric tests Mann-Whitney and Kruskal-Wallis test used to identify the existence of a significant difference in structural framework duration between independent groups of construction method. Parametric tests for comparison of group were not used due to unbalanced sample sizes in each group.

### **3.7.3 Hypotheses testing**

This study made two tentative predictions about existence of significant linear relationship among the response variable (structural works construction duration) and the two groups

of independent variables (project scope related and quantity of work items). using a simple linear regression model as in equation the null and alternative hypothesis are shown in standard equation format below.

$$Y = \beta_0 + \beta_1 x + \varepsilon \quad (24)$$

Where  $Y$ - is structural works construction duration,  $\beta_0$  is the y-intercept while  $\beta_1$  is the slope which indicates the amount of variation in  $Y$  for a unit increase in the independent variable. The predictor variables used in the hypotheses are project scope related and quantity of work item factors. And  $\varepsilon$  is the error term between the predicted and the actual value of the response variable  $Y$ .

Form the regression equation the hypothesis interest lies on the value of the slope coefficient or  $\beta_1$  term, and the hypothesis is set mathematically as follows;

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

The null hypothesis assumes that there is no significant linear relationship between the response and predictor variables. A two sided hypothesis were used because, any association in either positive or negative direction is important. The  $p$  values used to test whether each regression coefficient is significantly different from zero at a significance level of 0.05.

### 3.7.4 Validity of results

Predictive models based on linear regression analysis need to fulfill basic assumptions after they are developed, the checking of these assumptions referred in regression books as “*regression diagnostics*” (e.g., (Rawlings, et al., 1998; Chatterjee & Hadi, 2006; Sarstedt & Mooi, 2014)). These assumptions are about regression residuals (errors in the regression model) and they are four which require the errors to be:

- Normally distributed,
- Independence of errors (lack of autocorrelation),
- Have constant variance commonly known as Homoscedasticity,
- Have a mean of zero,

Apart from the four above assumptions regarding residuals a basic underlining assumption of linear regression is the fact that there exists a linear relationship between the response variable and selected predictor variable. this assumption is termed as linearity assumption.

When using multiple regression analysis, the multicollinearity assumption that requires the absence of high correlation among predictor variables need to be checked in addition.

It is suggested to assume a model to be valid during a mild violation of the four assumptions of residuals and make necessary mathematical transformation of variables during clear violations (Pardoe, 2012 ).

Table 3-6 below show the assumptions tested together with their respective indicators and accepted values.

**Table 3-6 Regression diagnostics used for models, slightly modified from (Guerrero, 2016)**

Assumption/problem to be tested	Indicator	Accepted values
Linearity	Scatter plot	
Residual analysis		
Independence	Durbin-Watson	1.5<d<2.5
Constant variance	Scatter plot	
Normality	Histogram Kolmogorov-Smirnov	
Multi-collinearity	Variance Inflation Factor (VIF)	≤10

Data values which are found at extreme distance from other set of data are known as outliers, the identification of these points is essential since they have significant influence on the distribution and mean value of the variable also the final result of the analysis (Barton & Peat, 2014 ). Cook’s distance with acceptance value of  $\leq 1$  used to detect outliers. Decision regarding outliers include, rejection of the data value, retain it and use non-parametric test or transformation of variable.

Snee, (1977) provided four procedures for checking the validity of a regression model:

- 1) Comparison of model predictions and coefficients with physical theory,
- 2) Collection of new data to check model predictions,
- 3) Comparison of results with theoretical models and simulated data,
- 4) Reservation of a portion of the available data to obtain an independent measure of the model prediction accuracy.

Procedures indicated in number 1 and 4 performed for conforming the validity of models developed. The physical theory regarding duration is that the construction duration calculated using models should be non-zero positive number with acceptable minimum value. While the fourth procedure was used by splitting the data set into two; model development and validation set. This procedure is also referred as cross-validation technique (Snee, 1977). A suggestion from Barton & Peat, (2014) was used to split the data into 80% for model building and 20% for validation this splitting was done prior to analysis by randomly selecting validation data from the total sample data.

### **3.8 Survey Distribution and Response Statistics**

During the planning stage of data collection stage and sampling of participants 60 companies were shortlisted for the survey. 46 companies were contacted both through telephone and in person to identify their willingness to participate on the survey. 32 real-estate development companies accepted the request. A survey package of 16 pages; 13 pages' data sheets separated into three groups of building projects (5 pages each for apartment and villa, and 3 pages for commercial buildings), 2 pages' instruction and definitions of survey terms and a 1-page letter of cooperation from the university were delivered.

The data collection was conducted from the end of May to the beginning of August. The Corona virus outbreak together with other political incidents in the capital city Addis Ababa greatly impacted the data collection. The incidents lead to a collection of lower number of data sets which reshaped the study. Only 14 real-estate companies returned the questionnaires. Table 3-7 below shows the numbers related with the questionnaire survey response rate. The returned data sheets of commercial and villa buildings were excluded from further analysis due to their small amount. The effective returned samples of 41 apartment projects were used for further analysis.

**Table 3-7 Summary of questionnaire survey response rate**

<b>Project category</b>	<b>Structural framework construction duration survey questionnaire</b>					<b>Remark</b>
	Planned	Delivered	Returned	Invalid (Rejected)	Effective Response rate	
Apartment	300	160	43	2	25%	Used for analysis
Commercial	180	96	3	0	3%	Not used
Villa	300	160	9	0	5%	Not used
<b>Total</b>	<b>780</b>	<b>416</b>	<b>55</b>	<b>2</b>	<b>12%</b>	

The next chapter present the analysis results and discussions of the research questions and hypotheses together with the development of duration prediction models.

## CHAPTER 4 RESULTS AND DISCUSSIONS

### 4.1 Introduction to Results and Discussions Section

This chapter presents the answers for the three research questions and results of the two hypotheses. The research questions identified as RQ-1, RQ-2, and RQ-3 and hypotheses as H1 and H2 are presented here below.

RQ-1. What is the kind and degree of relationship among the dependent structural framework construction duration variables (i.e. total, substructure & super structure construction durations) and each of independent selected project scope parameters (i.e. Gross and typical floor area, total number of floors above and below ground, total and typical building height)?

RQ-2. What is the kind and degree of association among the dependent structural framework construction duration variables (i.e. total, substructure & super structure construction durations) and each of independent selected quantities of work items (i.e. quantity of concrete, excavation, formwork and rebar)?

RQ-3. Is there a difference in structural framework construction durations by selected construction method (as measured in foundation type, slab type, use of concrete pump and tower crane, concrete mixing method and type of shoring and scaffolding system used for formwork)?

H1: there is a significant linear relation between structural works duration and selected project scope factors (i.e. Gross and typical floor area, total number of floors above and below ground, total and typical building height.)

H2: there is a significant linear relation between structural works duration and selected quantity of work elements (i.e. volume of concrete, excavation, quantity of formwork and rebar).

This chapter composed of eight sections starts with statistical description of the data set followed by presenting the results and discussions of the first two research question and hypotheses. Prior to the results and discussions of the third research question, analysis results and discussions on structural framework duration prediction models presented.

## 4.2 General Description of Data Sets

### 4.2.1 Data set

The data set composed of samples of previously constructed 41 apartment projects whose information was gathered using structured questionnaire survey. This data set divided into model building and validation data set for the purpose of addressing the first two research questions and hypotheses while the third research question used the complete data set without such division. Numerical data can be obtained in the appendices section of the thesis.

### 4.2.2 Model building data set

The literature review section in chapter two indicated the importance of cross validation of a predictive model. For this purpose, the apartment data set were separated into model building and validation data set. Approximately 80% of the total data set which is 33 samples were used for generating predictive models of structural framework construction duration. Table 4-1, Table 4-2, and Table 4-3 below provide the descriptive statics of the model building data set.

**Table 4-1 Descriptive summary of dependent variables of model building data set**

Dependent variable	Unit	N (Sample size)	Minimum	Maximum	Mean	Std. Deviation
<i>D_Total</i>	days	33	130.0	2,100.0	799.909	527.489
<i>D_Super</i>	days	33	90.0	1,200.0	524.394	324.660
<i>D_Sub</i>	days	33	30.0	900.0	273.242	268.854

**Table 4-2 Descriptive summary of project scope variables of model building data set**

Independent variable	Unit	N (Sample size)	Minimum	Maximum	Mean	Std. Deviation
<i>T_Floors</i>	number	33	6.0	20.0	11.424	3.428
<i>B_Floors</i>	number	33	.0	2.0	1.242	0.751
<i>A_Floors</i>	number	33	4.0	19.0	9.212	3.257
<i>T_Height</i>	m	33	14.0	76.0	38.652	13.341
<i>Ty_height</i>	m	33	3.0	3.8	3.103	0.192
<i>T_GFA</i>	m <sup>2</sup>	33	1,193.0	33,072.0	10,635.091	8,043.517
<i>Ty_area</i>	m <sup>2</sup>	33	200.0	2,100.0	874.485	524.819

**Table 4-3 Descriptive summary of quantity of works variables of model building data set**

Independent variable	Unit	N (Sample size)	Minimum	Maximum	Mean	Std. Deviation
<i>Q_excav</i>	m <sup>3</sup>	33	120.00	14,720.00	5,387.654	4,243.265
<i>Q_form</i>	m <sup>2</sup>	33	2,700.00	32,000.00	15,032.033	9,192.033
<i>Q_bar</i>	Kg	33	25,737.40	780,727.00	321,061.605	25,3775.432
<i>Q_conc</i>	m <sup>3</sup>	33	936.00	14,847.00	5,481.439	3,665.095

#### 4.2.3 Model validation data set

Validating predictive models using independent data set from that used for generating predictive models suggested by several authors in previous related studies. Approximately 20% of the data set used for validation purpose. Eight samples were randomly excluded from the entire data set using Excels ‘*Rand*’ function. Table 4-4 below shows a compacted descriptive summary of all quantitative variables of the validation data set.

**Table 4-4 Descriptive summary of validation data set**

Variable	Unit	N (Sample size)	Minimum	Maximum	Mean	Std. Deviation
<i>D_Total</i>	days	8	182.0	1601.0	849.750	503.1685
<i>D_Super</i>	days	8	110.0	1055.0	595.750	344.9641
<i>D_Sub</i>	days	8	60.0	700.0	252.500	249.4566
<i>T_Floors</i>	number	8	7.0	18.0	12.250	3.6936
<i>B_Floors</i>	number	8	1.0	2.0	1.375	0.5175
<i>A_Floors</i>	number	8	4.0	16.0	9.875	3.943
<i>T_Height</i>	m	8	29.0	62.0	43.175	13.076
<i>Ty_height</i>	m	8	3.0	3.9	3.113	0.3182
<i>T_GFA</i>	m <sup>2</sup>	8	6,096.0	31,518.0	14,767.875	7,789.389
<i>Ty_area</i>	m <sup>2</sup>	8	381.0	1,830.0	1,233.625	539.724
<i>Q_excav</i>	m <sup>3</sup>	8	2,800.00	9,402.00	4,569.500	2,363.7943
<i>Q_form</i>	m <sup>2</sup>	8	3,100.00	25,696.00	15,318.750	9,395.2190
<i>Q_bar</i>	Kg	8	31,500.00	655,256.00	309,597.625	272,186.0683
<i>Q_conc</i>	m <sup>3</sup>	8	936.00	11,926.00	4,469.125	3,927.9651

## 4.3 Correlation Analysis, Results and Discussions

### 4.3.1 Structural works duration correlation with project scope factors

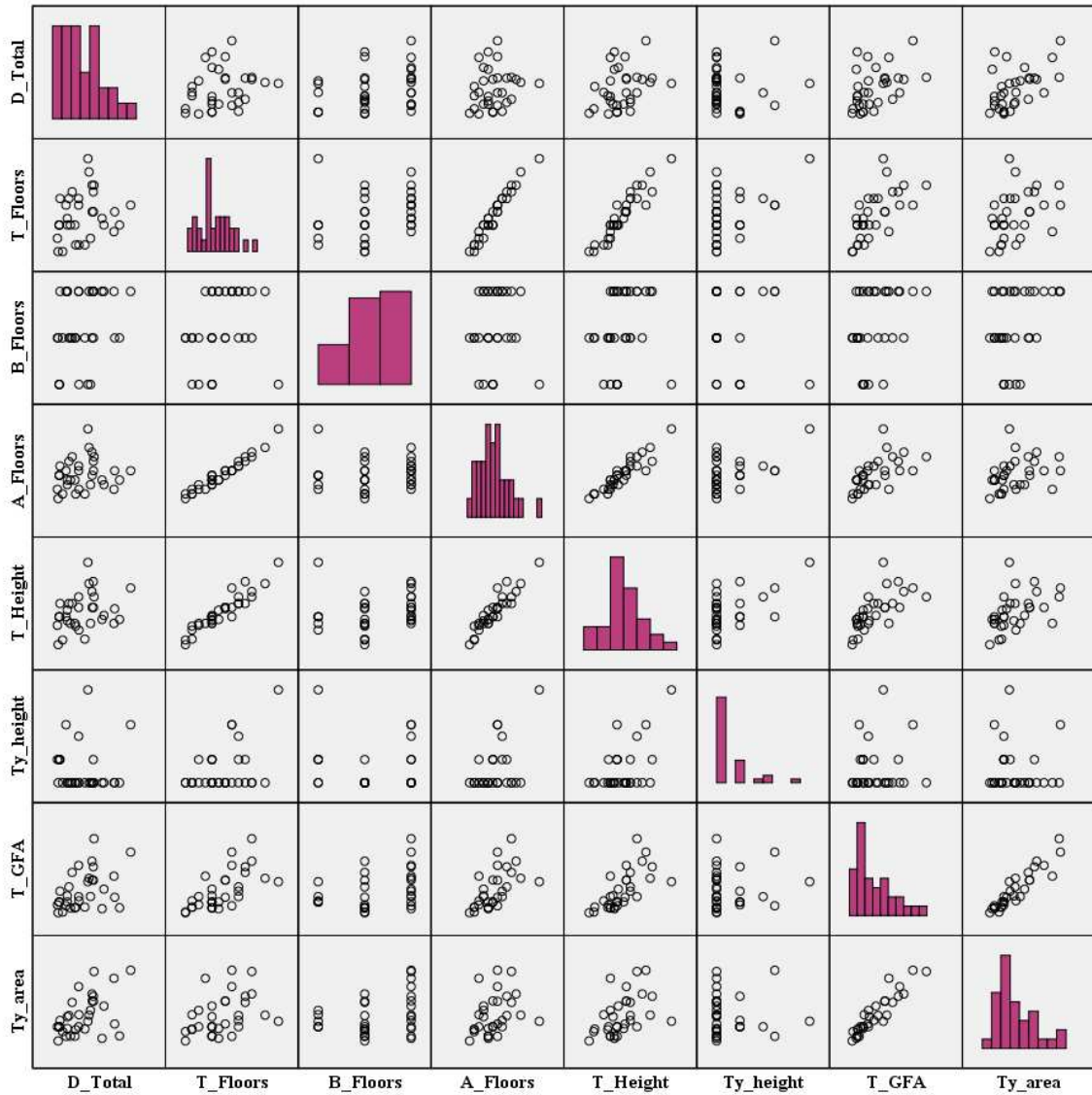
The first set of research question aimed to identify the possible association among the three dependent variables of interest (i.e. total structural works, superstructure and substructure works durations) and proposed set of independent variables. The associations were tested using Pearson's Correlation Coefficient ( $r$ ). Before commencing with identification of significant associations among dependent and independent variables tests were conducted for possible violation of assumptions.

#### 4.3.1.1 Statistical analysis procedure

Bivariate linear associations between dependent variable and sets of independent variables analyzed using Pearson's product moment correlation. The preliminary analysis involved the checking of possible violation of the following three assumptions:

1. **Continuous data for each variable:** this assumption requires the variables to be of interval or ratio level discarding categorical level data. A brief discussion of the variables used in the study is outlined in the methodology section.
2. **Independent observations:** samples in the data set should be independent of one another, i.e. the value of one sample should not affect the other. This assumption is justified as samples are unique projects among themselves.
3. **Linearly related variables:** the existence of a linear association assessed using scatter plot during violations of the assumption transformation of variables used.
4. **Normally distributed variables:** histogram of variables was used to determine possible deviations from normal distribution. In most cases variables found to be positively skewed.

Figure 4-1 below displays a sample output of scatter plot matrix together with histogram on diagonal for total structural frame work construction duration and project scope variables. From the scatter plots their seems a positive linear relationship only between the two project scope variables  $T\_GFA$  &  $Ty\_area$  with  $D\_Total$ . The other scatter plots suggest violations of linearity assumption.



**Figure 4-1 Matrix scatter plot together with histogram *D\_Total* variable**

Violations of linearity and normality assumptions are adjusted with transformation of the observed data. Therefore, bivariate correlations were conducted using observed data in its natural form and with selected mathematical transformation.

**4.3.1.2 Bivariate correlations without transformations**

Table 4-5 below provides the Pearson’s correlation coefficient among dependent variables and the seven-project scope independent variables before data transformation.

**Table 4-5 Pearson's correlations between dependent variables & project scope factors**

		<i>T_Floors</i>	<i>B_Floor</i>	<i>A_Floors</i>	<i>T_Height</i>	<i>Ty_height</i>	<i>T_GFA</i>	<i>Ty_area</i>
<i>D_Total</i>	<i>r</i>	0.223	0.303	0.155	0.325	-0.017	0.456**	0.510**
	<i>Sig.</i>	0.212	0.086	0.388	0.065	0.923	0.008	0.002
<i>D_Super</i>	<i>r</i>	0.363*	0.330	0.297	0.464**	0.092	0.537**	0.533**
	<i>Sig.</i>	0.038	0.061	0.094	0.006	0.611	0.001	0.001
<i>D_Sub</i>	<i>r</i>	-0.007	0.196	-0.060	0.070	-0.147	0.244	0.356*
	<i>Sig.</i>	0.970	0.275	0.741	0.698	0.416	0.172	0.042

\*. The correlation is significant at the 0.05 level. - \*\*. The correlation is significant at the 0.01 level.

The correlation matrix above showed only seven statistically significant correlations from the possible 21. This is attributed to the violation of the two core assumptions of correlation statistics (i.e. linearity and normality assumptions). Four recommended mathematical data transformations were used and the resulting correlations investigated.

#### 4.3.1.3 Bivariate correlations with variable transformations

The necessity for transformation of variables arise from the violation of assumptions of linearity between dependent variable and independent variables and normal distribution of variables. Logarithmic, square root and inverse mathematical transformations were conducted on the dependent variables and improvement in the correlation coefficients magnitude were obtained. Table 4-6 shows the output of correlation analysis between transformed dependent variable and original independent project scope variables.

**Table 4-6 Pearson's correlation of transformed *D\_Total* with project scope factors**

<i>D_Total</i>		<i>T_Floors</i>	<i>B_Floors</i>	<i>A_Floors</i>	<i>T_Height</i>	<i>Ty_heig</i>	<i>T_GFA</i>	<i>Ty_area</i>
<i>Observed</i>	<i>r</i>	0.223	0.303	0.155	0.325	-0.017	0.456**	0.510**
	<i>Sig.</i>	0.212	0.086	0.388	0.065	0.923	0.008	0.002
<i>INV</i>	<i>r</i>	<b>-0.351*</b>	-0.336	-0.285	-0.376*	0.121	-0.438*	-0.439*
	<i>Sig.</i>	<b>0.045</b>	0.056	0.108	0.031	0.502	0.011	0.011
<i>LOG</i>	<i>r</i>	0.313	0.329	0.243	<b>0.378*</b>	-0.086	<b>0.481**</b>	0.505**
	<i>Sig.</i>	0.076	0.061	0.173	<b>0.030</b>	0.634	<b>0.005</b>	0.003
<i>SQRT</i>	<i>r</i>	0.272	0.318	0.203	0.357*	-0.054	0.477**	<b>0.517**</b>
	<i>Sig.</i>	0.126	0.071	0.258	0.041	0.764	0.005	<b>0.002</b>

\*. The correlation is significant at the 0.05 level. - \*\*. The correlation is significant at the 0.01 level.

Higher magnitudes of correlations observed with transformed data of the dependent variable. Number of floors above and below ground and typical floor height showed no significant linear association with total structural framework duration even after data transformation.

The correlation values of the transformed data of the superstructure duration summarized below together with the observed ones in Table 4-7.

**Table 4-7 Pearson’s correlation of transformed *D\_Super* with project scope factors**

<i>D_Super</i>		<i>T_Floors</i>	<i>B_Floors</i>	<i>A_Floors</i>	<i>T_Height</i>	<i>Ty_heig ht</i>	<i>T_GFA</i>	<i>Ty_area</i>
<i>Observed</i>	<i>r</i>	0.363*	0.330	0.297	0.464**	0.092	<b>0.537**</b>	<b>0.533**</b>
	<i>Sig.</i>	0.038	0.061	0.094	0.006	0.611	<b>0.001</b>	<b>0.001</b>
<i>INV</i>	<i>r</i>	-0.424*	-0.343	<b>-0.361*</b>	-0.450**	0.055	-0.451**	-0.422*
	<i>Sig.</i>	0.014	0.051	<b>0.039</b>	0.009	0.760	0.008	0.014
<i>LOG</i>	<i>r</i>	<b>0.411*</b>	<b>0.359*</b>	0.341	<b>0.473**</b>	0.000	0.514**	0.492**
	<i>Sig.</i>	<b>0.017</b>	<b>0.040</b>	0.052	<b>0.005</b>	1.000	0.002	0.004
<i>SQRT</i>	<i>r</i>	0.390*	0.348*	0.321	0.472**	0.045	0.531**	0.517**
	<i>Sig.</i>	0.025	0.048	0.069	0.006	0.805	0.001	0.002

\*. The correlation is significant at the 0.05 level. - \*\*. The correlation is significant at the 0.01 level.

Table 4-8 below summarizes the correlations among transformed data of the substructure duration and project scope variables.

**Table 4-8 Pearson’s correlation of transformed *D\_Sub* with project scope factors**

<i>D_Sub</i>		<i>T_Floors</i>	<i>B_Floors</i>	<i>A_Floors</i>	<i>T_Height</i>	<i>Ty_height</i>	<i>T_GFA</i>	<i>Ty_area</i>
<i>Observed</i>	<i>r</i>	-0.007	0.196	-0.060	0.070	-0.147	0.244	0.356*
	<i>Sig.</i>	0.970	0.275	0.741	0.698	0.416	0.172	0.042
<i>INV</i>	<i>r</i>	-0.117	<b>-0.488**</b>	-0.006	-0.084	0.328	-0.316	-0.360*
	<i>Sig.</i>	0.515	<b>0.004</b>	0.976	0.643	0.063	0.073	0.040
<i>LOG</i>	<i>r</i>	0.061	0.320	-0.018	0.081	-0.276	0.315	<b>0.403*</b>
	<i>Sig.</i>	0.736	0.069	0.920	0.652	0.120	0.074	<b>0.020</b>
<i>SQRT</i>	<i>r</i>	0.024	0.242	-0.039	0.075	-0.212	0.281	0.385*
	<i>Sig.</i>	0.897	0.175	0.828	0.677	0.236	0.113	0.027

\*. The correlation is significant at the 0.05 level. - \*\*. The correlation is significant at the 0.01 level.

The inverse transformation of the substructure duration creates a statistically significant negative moderate association with the project scope variable number of floors below ground.

Overall the data transformation the dependent variables entails two major outcomes; either increased a statistically significant correlation magnitude or made new significant correlations.

#### 4.3.1.4 Discussions on correlation results of project scope factors

A correlation analysis assesses the kind and degree of association among the variables of interest. Three kinds of relationship do exist, perfectly linear, positive, and negative. Different authors suggest different recommendation for the degree of association among variables observed from magnitude of the correlation coefficient. Table 4-9 below show such a suggestion which was used for analyzing the magnitudes of correlation obtained between the dependent variables and selected project scope independent variables in Table 4-5.

**Table 4-9 Suggestion on strength of linear correlation adopted from (Jackson, 2009)**

Value of $r$	Degree (strength) of association
$\pm 0.7 \leq r \leq 1$	Strong
$\pm 0.3 \leq r \leq 0.69$	Moderate
$\pm 0.0 \leq r \leq 0.29$	None (0) to Weak

Typical floor area ( $Ty\_area$ ) showed statistically significant moderate positive correlation with all the three dependent variables. Except with the substructure duration ( $D\_Sub$ ), total gross floor area ( $T\_GFA$ ) also showed similar significant correlation. Table 4-5 also indicates a significant moderate positive correlation among super structure duration ( $D\_Super$ ) and two of the project scope variables: i.e. Total number of floors ( $T\_Floors$ ) and total building height ( $T\_Height$ ).

Prior studies indicated strong correlations among project scope variables and construction duration of buildings. But here only number of variables showed moderate associations with structural framework construction duration variables. This is due to the violation of the normal distribution assumption of the correlation analysis. Transformation of variables resulted in either increased magnitude of association or in new statistically significant association which were not possible on the observed data.

### 4.3.2 Structural works duration correlation with quantity of work items

The second research question intended to identify the degree of association among the three dependent variables (i.e.  $D_{Total}$ ,  $D_{Super}$ ,  $D_{Sub}$ ) and selected quantity of works independent variables. Quantity of concrete, excavation, formwork, and reinforcement bar are the selected independent variables whose linear association was analyzed. Analysis and results are presented in similar fashion as research question 1 with less detail in this section.

A sample output of this preliminary investigation for possible violations of linearity and normality assumption given below in Figure 4-2 for superstructure construction duration. Correlation analysis was conducted using observed data and transformed dependent variable.



Figure 4-2 Matrix scatter plot together with histogram  $D_{Super}$  variable

#### 4.3.2.1 Bivariate correlations without transformations

Table 4-10 below shows the Pearson's correlation coefficient among the three dependent variables of structural framework construction duration and the quantity of works variables.

**Table 4-10 Pearson's correlations between dependent variables & quantity of work items**

		<i>Q_excav</i>	<i>Q_form</i>	<i>Q_bar</i>	<i>Q_conc</i>
<i>D_Total</i>	<i>r</i>	0.537**	0.412*	0.591**	0.328
	<i>Sig.</i>	0.001	0.017	0.000	0.063
<i>D_Super</i>	<i>r</i>	0.504**	0.362*	0.481**	0.351*
	<i>Sig.</i>	0.003	0.038	0.005	0.046
<i>D_Sub</i>	<i>r</i>	0.441*	0.365*	0.574**	0.212
	<i>Sig.</i>	0.010	0.037	0.000	0.236

\*. The correlation is significant at the 0.05 level. - \*\*. The correlation is significant at the 0.01 level.

Statistically significant moderate positive correlations obtained between quantity of excavation, formwork and reinforcement bar with all the three dependent variables. While quantity of concrete heled similar correlation only with superstructure duration.

#### 4.3.2.2 Bivariate correlations with variable transformations

Transformation of the dependent variables were done not only because of violation of assumptions of linearity and normality but to check also for improved magnitude of correlation. Table 4-11 below shows the results of correlation analysis among transformed total structural works duration and quantity of works selected variables.

**Table 4-11 Pearson's correlation of transformed *D\_Total* with quantity of work items**

<i>D_Total</i>		<i>Q_excav</i>	<i>Q_form</i>	<i>Q_bar</i>	<i>Q_conc</i>
<i>Observed</i>	<i>r</i>	0.537**	0.412*	<b>0.591**</b>	0.328
	<i>Sig.</i>	0.001	0.017	<b>0.000</b>	0.063
<i>INV</i>	<i>r</i>	-0.548**	<b>-0.539**</b>	-0.511**	-0.338
	<i>Sig.</i>	0.001	<b>0.001</b>	0.002	0.055
<i>LOG</i>	<i>r</i>	<b>0.555**</b>	0.522**	0.576**	<b>0.363*</b>
	<i>Sig.</i>	<b>0.001</b>	0.002	0.000	<b>0.038</b>
<i>SQRT</i>	<i>r</i>	0.548**	0.475**	<b>0.591**</b>	0.354*
	<i>Sig.</i>	0.001	0.005	<b>0.000</b>	0.043

\*. The correlation is significant at the 0.05 level. - \*\*. The correlation is significant at the 0.01 level.

Logarithmic transformation of data held a significant positive moderate association with quantity of concrete variable.

Table 4-12 present the correlations obtained by transforming only the dependent variable superstructure duration.

**Table 4-12 Pearson's correlation of transformed  $D\_Super$  with quantity of work items**

$D\_Super$		$Q\_exc$	$Q\_form$	$Q\_bar$	$Q\_conc$
<i>Observed</i>	<i>r</i>	0.504**	0.362*	0.481**	<b>0.351*</b>
	<i>Sig.</i>	0.003	0.038	0.005	<b>0.046</b>
<i>INV</i>	<i>r</i>	-0.526**	<b>-0.412*</b>	-0.478**	-0.279
	<i>Sig.</i>	0.002	<b>0.017</b>	0.005	0.116
<i>LOG</i>	<i>r</i>	<b>0.535**</b>	0.410*	<b>0.507**</b>	0.331
	<i>Sig.</i>	<b>0.001</b>	0.018	<b>0.003</b>	0.060
<i>SQRT</i>	<i>r</i>	0.522**	0.390*	0.499**	0.346*
	<i>Sig.</i>	0.002	0.025	0.003	0.049

\*. The correlation is significant at the 0.05 level. - \*\*. The correlation is significant at the 0.01 level.

The transformed dependent variable resulted in an increased degree of association among the three quantity of work variables (i.e. quantity of excavation, formwork, and rebar).

Table 4-13 below summarises the correlation coefficients resulted after transformation of variables. Unlike the dependent variables  $D\_Total$  and  $D\_Super$  here increased correlations were obtained when both the dependent and independent variables were transformed together.

**Table 4-13 Pearson's correlation of transformed  $D\_Sub$  with quantity of work items**

$D\_Sub$		$Q\_exc$	$Q\_form$	$Q\_bar$	$Q\_conc$
<i>Observed</i>	<i>r</i>	0.441*	0.365*	0.574**	0.212
	<i>Sig.</i>	0.010	0.037	0.000	0.236
<i>INV</i>	<i>r</i>	<b>0.648**</b>	<b>0.736**</b>	<b>0.716**</b>	0.285
	<i>Sig.</i>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	0.107
<i>LOG</i>	<i>r</i>	0.486**	0.555**	0.655**	0.319
	<i>Sig.</i>	0.004	0.001	0.000	0.070
<i>SQRT</i>	<i>r</i>	0.445**	0.455**	0.608**	0.275
	<i>Sig.</i>	0.009	0.008	0.000	0.121

\*. The correlation is significant at the 0.05 level. - \*\*. The correlation is significant at the 0.01 level.

No significant correlation was obtained between quantity of concrete and substructure duration.

#### **4.3.2.3 Discussions on correlation results of quantity of work items**

Statistically significant moderate positive bivariate correlations obtained between three of the selected quantity of work item variables (i.e.  $Q_{exc}$ ,  $Q_{form}$ , and  $Q_{bar}$ ) and dependent construction duration (total, substructure, and superstructure) variables. Quantity of concrete showed a positive moderate correlation only with superstructure duration. All the correlations observed was positive indicating an increase in quantity of work items leading to increased durations.

Transformation of the two dependent variables while keeping the independent variables original form lead to higher magnitudes of correlation. While such similar results obtained by transforming both the variables in case of substructure duration. The quantity of concrete showed no relation with substructure duration in all transformations and in observed data scenario. The inverse transformation of  $D_{Sub}$  and  $Q_{form}$  resulted in the highest correlation coefficient  $Q_{bar}$  also showed a strong positive association due to the transformation.

The analysis and results of research question 1 and 2 is a step forward for answering the preliminary research hypotheses set in the research.

## **4.4 Analysis, Results and Discussions of Hypotheses**

### **4.4.1 Hypothesis of linear association with project scope factors**

The first research hypothesis states that: there is a significant linear relation between structural works duration and selected project scope factors (i.e. Gross and typical floor area, total number of floors above and below ground, total and typical building height.)

#### **4.4.1.1 Statistical procedure to test hypothesis 1 (H1)**

Simple linear regression of the form in equation 25 used to test the existence of linear relation.

$$Y_i = \beta_0 + \beta_1 x_j \quad (25)$$

Where:  $Y_i$  = structural frame work durations in days ( $D_{Total}$ ,  $D_{Super}$ ,  $D_{Sub}$ )

$X_j$  = selected project scope variables ( $T\_Floors$ ,  $B\_Floors$ ,  $A\_Floors$ ,  $T\_Height$ ,  $Ty\_height$ ,  $T\_GFA$ ,  $Ty\_area$ )

$\beta_0$  = y intercept and  $\beta_1$  slope of the linear equation

$i=1,2,3$  and  $j=1,2,3\dots7$

A two-tailed hypothesis test for the slope of the linear equation stated as;

Null hypothesis NH:  $\beta_1 = 0$  and the Alternative hypothesis AH:  $\beta_1 \neq 0$

A statistical significance level of 5% used to reject or accept the null hypothesis.

#### 4.4.1.2 Results of Hypothesis 1 (H1)

- **Total structural framework duration with project scope variables**

The results of simple linear regression analysis between project scope and total duration presented in Table 1 14.

**Table 4-14 Statistical results of  $D\_Total$  for Hypothesis 1 (H1)**

$D\_Total$ Model		Unstandardized		t	Sig.	95.0% Confidence Interval	
		Coefficients				for B	
		B	Std. Error			Lower Bound	Upper Bound
1	(Constant)	407.772	320.922	1.271	0.213	-246.753	1062.298
	$T\_Floors$	34.325	26.940	1.274	0.212	-20.619	89.269
2	(Constant)	535.354	173.764	3.081	0.004	180.960	889.748
	$B\_Floors$	212.935	120.169	1.772	0.086	-32.151	458.020
3	(Constant)	568.392	280.280	2.028	0.051	-3.243	1140.027
	$A\_Floors$	25.132	28.733	.875	0.388	-33.470	83.734
4	(Constant)	303.253	274.162	1.106	0.277	-255.904	862.410
	$T\_Height$	12.850	6.716	1.913	0.065	-0.847	26.547
5	(Constant)	947.773	1527.289	.621	0.539	-2167.153	4062.700
	$Ty\_height$	-47.652	491.274	-.097	0.923	-1049.611	954.308
6	(Constant)	482.143	139.035	3.468	0.002	198.579	765.708
	$T\_GFA$	<b>0.030</b>	<b>0.010</b>	<b>2.850</b>	<b>0.008</b>	<b>0.008</b>	<b>0.051</b>
7	(Constant)	351.377	157.694	2.228	0.033	29.758	672.996
	$Ty\_area$	<b>0.513</b>	<b>0.155</b>	<b>3.304</b>	<b>0.002</b>	<b>0.196</b>	<b>0.830</b>

Statistically significant positive linear relations observed for total gross floor area ( $T\_GFA$ ) and typical floor area ( $Ty\_area$ ) project scope variables with total duration. A p-value of the t-statistics for the slope coefficients are 0.008 for  $T\_GFA$  and 0.002 for  $Ty\_area$  both less than the 0.05 significance level therefore the null hypothesis of no significant linear relation rejected for  $D\_Total$  and the alternative hypothesis accepted.

- **Superstructure duration with project scope variables**

The results of simple linear regression analysis between project scope and superstructure duration presented in Table 4-15.

**Table 4-15 Statistical results of  $D\_Super$  for Hypothesis 1 (H1)**

$D\_Super$ Model		Unstandardized Coefficients		t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error			Lower Bound	Upper Bound
<b>1</b>	(Constant)	131.218	188.775	0.695	0.492	-253.792	516.227
	<b><math>T\_Floors</math></b>	<b>34.416</b>	<b>15.847</b>	<b>2.172</b>	<b>0.038</b>	<b>2.096</b>	<b>66.736</b>
2	(Constant)	347.196	105.946	3.277	0.003	131.118	563.274
	$B\_Floors$	142.622	73.268	1.947	0.061	-6.809	292.054
3	(Constant)	251.916	166.757	1.511	0.141	-88.187	592.019
	$A\_Floors$	29.578	17.095	1.730	0.094	-5.288	64.444
4	(Constant)	87.622	158.024	0.554	0.583	-234.670	409.913
	<b><math>T\_Height</math></b>	<b>11.300</b>	<b>3.871</b>	<b>2.919</b>	<b>0.006</b>	<b>3.405</b>	<b>19.195</b>
5	(Constant)	43.796	936.176	0.047	0.963	-1865.549	1953.140
	$Ty\_height$	154.880	301.134	0.514	0.611	-459.287	769.048
6	(Constant)	293.769	81.079	3.623	0.001	128.408	459.131
	<b><math>T\_GFA</math></b>	<b>0.022</b>	<b>0.006</b>	<b>3.547</b>	<b>0.001</b>	<b>0.009</b>	<b>0.034</b>
7	(Constant)	235.812	95.460	2.470	0.019	41.120	430.504
	<b><math>Ty\_area</math></b>	<b>0.330</b>	<b>0.094</b>	<b>3.512</b>	<b>0.001</b>	<b>0.138</b>	<b>0.522</b>

Project scope variables;  $T\_Floors$ ,  $T\_Height$ ,  $T\_GFA$ , and  $Ty\_area$  showed significant positive linear association with the dependent variable  $D\_Super$ . Therefore, the null hypothesis of no significant linear relation was rejected for  $D\_Super$  and the alternative hypothesis accepted.

- **Substructure duration with project scope variables**

Similar hypothesis test was conducted for the third dependent variable substructure duration for identification of significant linear association with selected project scope variables. Results of the simple linear regression are presented in Table 4-16 below.

**Table 4-16 Statistical results of  $D_{Sub}$  for Hypothesis 1 (H1)**

$D_{Sub}$ Model		Unstandardized		t	Sig.	95.0% Confidence Interval	
		Coefficients				for B	
		B	Std. Error			Lower Bound	Upper Bound
1	(Constant)	279.330	167.795	1.665	0.106	-62.889	621.550
	$T_{Floors}$	-0.533	14.086	-0.038	0.970	-29.260	28.195
2	(Constant)	186.285	91.147	2.044	0.050	0.389	372.182
	$B_{Floors}$	69.990	63.034	1.110	0.275	-58.569	198.549
3	(Constant)	318.628	144.349	2.207	0.035	24.225	613.031
	$A_{Floors}$	-4.927	14.798	-0.333	0.741	-35.108	25.254
4	(Constant)	218.540	147.393	1.483	0.148	-82.069	519.150
	$T_{Height}$	1.415	3.611	0.392	0.698	-5.948	8.779
5	(Constant)	907.111	770.158	1.178	0.248	-663.637	2477.859
	$Ty_{height}$	-204.274	247.732	-0.825	0.416	-709.527	300.979
6	(Constant)	186.648	77.209	2.417	0.022	29.179	344.117
	$T_{GFA}$	0.008	0.006	1.398	0.172	-0.004	0.020
7	(Constant)	113.636	87.327	1.301	0.203	-64.469	291.741
	$Ty_{area}$	<b>0.183</b>	<b>0.086</b>	<b>2.123</b>	<b>0.042</b>	<b>0.007</b>	<b>0.358</b>

A statistically significant positive linear relation observed between typical area ( $Ty_{area}$ ) and substructure duration ( $D_{Sub}$ ). A p-value of 0.042 for the t-statistics for the slope coefficient is less than the 0.05 significance level therefore, the null hypothesis of no significant linear relation rejected for  $D_{Sub}$  and the alternative hypothesis accepted.

#### 4.4.1.3 Discussions on results of Hypothesis 1 (H1)

A sample 33 apartment projects from model building data set was used to test the null hypothesis of no significant linear association between duration variables and project scope variables. Simple linear regression conducted on SPSS revealed existence of statistically significant positive linear relation. Project scope variable  $Ty_{area}$  showed significant positive relation with the three dependent variables (i.e.  $D_{Total}$ ,  $D_{Super}$ ,  $D_{Sub}$ ), while similar relation observed for  $T_{GFA}$  with the  $D_{Total}$  and  $D_{Super}$ . Further also significant positive linear association between project scope variables  $T_{Floors}$  and  $T_{Height}$  obtained with  $D_{Super}$ .

The p-values of the t-statistics of the slope coefficient for the above discussed project scope variables observed to be less than the significance value of the research 0.05 where the null hypothesis of no significant linear association rejected and the alternative hypothesis accepted.

#### 4.4.2 Hypothesis of linear association with quantity of work items

The second research hypothesis assumed that; there is a significant linear relation between structural works duration and selected quantity of work elements (i.e. volume of concrete, excavation, quantity of formwork and rebar).

##### 4.4.2.1 Statistical procedure to test Hypothesis 2 (H2)

Similar approach followed to test H2 as one used in the previous H1 here we assessed the interrelationship among quantity of work item variables with dependent duration variables.

Simple linear regression of the form in equation 28 used to test the existence of linear relation.

$$Y_i = \beta_0 + \beta_1 x_j \quad (26)$$

Where:  $Y_i$  = structural frame work durations ( $D\_Total$ ,  $D\_Super$ ,  $D\_Sub$ )

$X_j$  = selected quantity of work item variables ( $Q\_exc$ ,  $Q\_form$ ,  $Q\_bar$ ,  $Q\_conc$ )

$\beta_0$  = y intercept and  $\beta_1$  slope of the linear equation

$i=1,2,3$  and  $j=1,2,3,4$

A two-tailed hypothesis test for the slope of the linear equation stated as;

Null hypothesis NH:  $\beta_1 = 0$  and the Alternative hypothesis AH:  $\beta_1 \neq 0$

A statistical significance level of 5% used to reject or accept the null hypothesis.

##### 4.4.2.2 Results of Hypothesis 2 (H2)

- **Total structural framework duration with quantity of work item variables**

The results of simple linear regression analysis between quantity of work item variables and total structural frameworks construction duration presented in Table 4-17. Quantity of work items  $Q\_exc$ ,  $Q\_form$ , and  $Q\_bar$  showed statistically significant positive linear

association. Especially  $Q_{bar}$  showed highly statistical significant linear relation with  $D_{Total}$  ( $p < 0.001$ ). Therefore, the null hypothesis of no significant linear relation was rejected for  $D_{Total}$  and the alternative hypothesis accepted.

**Table 4-17 Statistical results of  $D_{Total}$  for Hypothesis 2 (H2)**

$D_{Total}$ Model		Unstandardized Coefficients		t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error			Lower Bound	Upper Bound
<b>1</b>	(Constant)	440.391	128.437	3.429	0.002	178.441	702.340
	$Q_{exc}$	<b>0.067</b>	<b>0.019</b>	<b>3.542</b>	<b>0.001</b>	<b>0.028</b>	<b>0.105</b>
<b>2</b>	(Constant)	444.816	164.818	2.699	0.011	108.667	780.964
	$Q_{form}$	<b>0.024</b>	<b>0.009</b>	<b>2.515</b>	<b>0.017</b>	<b>0.004</b>	<b>0.043</b>
<b>3</b>	(Constant)	405.653	122.548	3.310	0.002	155.715	655.592
	$Q_{bar}$	<b>0.001</b>	<b>0.000</b>	<b>4.077</b>	<b>0.000</b>	<b>0.001</b>	<b>0.002</b>
4	(Constant)	541.358	160.277	3.378	0.002	214.471	868.245
	$Q_{conc}$	0.047	0.024	1.931	0.063	-0.003	0.097

- **Superstructure duration with quantity of work item variables**

The results of simple linear regression analysis between quantity of work item variables and superstructure construction duration using the 33 project sample data presented in Table 4-18.

**Table 4-18 Statistical results of  $D_{Super}$  for Hypothesis 2 (H2)**

$D_{Super}$ Model		Unstandardized Coefficients		t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error			Lower Bound	Upper Bound
<b>1</b>	(Constant)	316.439	80.898	3.912	0.000	151.447	481.431
	$Q_{exc}$	<b>0.039</b>	<b>0.012</b>	<b>3.253</b>	<b>0.003</b>	<b>0.014</b>	<b>0.063</b>
<b>2</b>	(Constant)	332.203	103.762	3.202	0.003	120.579	543.827
	$Q_{form}$	<b>0.013</b>	<b>0.006</b>	<b>2.162</b>	<b>0.038</b>	<b>0.001</b>	<b>0.025</b>
<b>3</b>	(Constant)	326.807	81.957	3.988	0.000	159.655	493.960
	$Q_{bar}$	<b>0.001</b>	<b>0.000</b>	<b>3.055</b>	<b>0.005</b>	<b>0.000</b>	<b>0.001</b>
<b>4</b>	(Constant)	354.197	97.790	3.622	0.001	154.753	553.641
	$Q_{conc}$	<b>0.031</b>	<b>0.015</b>	<b>2.084</b>	<b>0.046</b>	<b>0.001</b>	<b>0.061</b>

A statistically significant positive linear association obtained between all the four quantity of work item variables and superstructure duration. The p-values for t-statistic of the slopes of quantity of work item variables are all less than the 0.05 significance level therefore, the null hypothesis of no significant linear association was rejected and the alternative hypothesis accepted.

- **Substructure duration with quantity of work item variables**

Similar to the total and superstructure dependent duration variables simple linear regression was conducted on substructure duration with quantity of work item variables and result presented in Table 4-19. Statistically significant positive linear relations obtained between  $Q_{exc}$ ,  $Q_{form}$ , and  $Q_{bar}$  with the dependent  $D_{Sub}$ . Highly statistically significant ( $P < 0.001$ ) relation obtained for  $Q_{bar}$ . Therefore, the null hypothesis of no significant linear relation rejected and the alternative hypothesis accepted.

**Table 4-19 Statistical results of  $D_{Sub}$  for Hypothesis 2 (H2)**

$D_{Sub}$ Model		Unstandardized		t	Sig.	95.0% Confidence Interval	
		Coefficients				for B	
		B	Std. Error			Lower Bound	Upper Bound
<b>1</b>	(Constant)	122.536	69.618	1.760	0.088	-19.451	264.523
	$Q_{exc}$	<b>0.028</b>	<b>0.010</b>	<b>2.740</b>	<b>0.010</b>	<b>0.007</b>	<b>0.049</b>
<b>2</b>	(Constant)	112.602	85.805	1.312	0.199	-62.398	287.602
	$Q_{form}$	<b>0.011</b>	<b>0.005</b>	<b>2.185</b>	<b>0.037</b>	<b>0.001</b>	<b>0.021</b>
<b>3</b>	(Constant)	78.073	63.403	1.231	0.227	-51.238	207.385
	$Q_{bar}$	<b>0.001</b>	<b>0.000</b>	<b>3.901</b>	<b>0.000</b>	<b>0.000</b>	<b>0.001</b>
4	(Constant)	187.952	84.499	2.224	0.034	15.615	360.289
	$Q_{conc}$	0.016	0.013	1.209	0.236	-0.011	0.042

#### 4.4.2.3 Discussions on results of Hypothesis 2 (H2)

A sample of 33 apartment projects was used to conduct simple linear regression and conduct hypothesis test on the slope coefficients of selected quantity of work item variables. Results of the t-statistic test on the slope coefficients were less than 0.05 for  $Q_{exc}$ ,  $Q_{form}$ , and  $Q_{bar}$  indicating existence of statistically significant linear relation with  $D_{Total}$ ,  $D_{Super}$ ,  $D_{Sub}$ .  $Q_{conc}$  showed a positive linear relation only with  $D_{Super}$ . Overall the null hypothesis was rejected for the three dependent variables and the alternative hypothesis of significant linear relation was accepted.

## 4.5 Structural Framework Construction Duration Prediction Models

One of the objectives of the research was to develop structural frameworks duration prediction models that can be used for future prediction purpose. Simple and multiple linear regression were carried out using SPSS and results presented with follow up discussions and interpretations.

### 4.5.1 Simple Linear Regression (SLR) Models

The bivariate correlations results suggested significant correlations among number of project scope and quantity of work independent variables and the three dependent variables of structural framework duration. Statistically significant independent variables from both correlation analysis and hypotheses test were used to generate simple linear regression (SLR) models.

- **Preliminary analysis**

Assumptions of simple linear regression presented in below.

Assumption/problem to be tested	Indicator
Linearity	Scatter plot
Residual analysis	
Independence	Durbin-Watson $1.5 < d < 2.5$
Constant variance	Scatter plot
Normality	Histogram

Among the four assumptions of linear regression only the assumption of linearity (linear relation between dependent and independent variable can be checked before conducting the analysis and this part was done during analysis of RQ-1 and RQ-2. The remaining three assumptions are dealt after SLR model development done and discussed as part of assessing the regression model for future use.

- **SLR models using project scope variables**

Project scope variables that were found to have a significant linear correlation during analysis of the first research question used to develop duration prediction models. Table 4-20 below shows the SLR models performance developed using project scope independent variables.

**Table 4-20 SLR models between dependent variables and project scope factors**

Dependent Variable	Independent Variable	R <sup>2</sup>	R <sup>2</sup> adjusted	F-value	Significance	Standard error of estimate	Durbin-Watson
<i>D_Total</i>	<i>T_GFA</i>	0.208	0.182	8.121	0.008	477.072	1.340
	<i>Ty_area</i>	0.260	0.237	10.916	0.002	460.893	1.275
<i>D_Super</i>	<i>T_Floors</i>	0.132	0.104	4.717	0.038	307.304	1.371
	<i>T_Height</i>	0.216	0.190	8.522	0.006	292.135	1.233
	<i>T_GFA</i>	0.289	0.266	12.579	0.001	278.206	1.328
	<i>Ty_area</i>	0.285	0.261	12.331	0.001	279.001	1.278
<i>D_Sub</i>	<i>Ty_area</i>	0.127	0.099	4.507	0.042	255.232	1.433

- **SLR models using quantity of work item variables**

Quantity of work item variables that were found to have a significant linear correlation during analysis of the second research question used to develop duration prediction models. Table 4-21 below shows the SLR models performance developed using project scope independent variables.

**Table 4-21 SLR models between dependent variables and quantity of work items**

Dependent Variable	Independent Variable	R <sup>2</sup>	R <sup>2</sup> adjusted	F-value	Significance	Standard error of estimate	Durbin-Watson
<i>D_Total</i>	<i>Q_excV</i>	0.288	0.265	12.548	0.001	452.170	1.655
	<i>Q_form</i>	0.169	0.143	6.325	0.017	488.416	1.292
	<i>Q_bar</i>	<b>0.349</b>	<b>0.328</b>	<b>16.621</b>	<b>0.000</b>	<b>432.405</b>	<b>1.933</b>
<i>D_Super</i>	<i>Q_excV</i>	<b>0.254</b>	<b>0.230</b>	<b>10.583</b>	<b>0.003</b>	<b>284.805</b>	<b>1.557</b>
	<i>Q_form</i>	0.131	0.103	4.675	0.038	307.4851	1.452
	<i>Q_bar</i>	0.231	0.207	9.333	0.005	289.182	1.704
	<i>Q_conc</i>	0.123	0.095	4.342	0.046	308.927	1.265
<i>D_Sub</i>	<i>Q_excV</i>	0.195	0.169	7.505	0.010	245.095	1.692
	<i>Q_form</i>	0.133	0.106	4.776	0.037	254.2715	1.238
	<i>Q_bar</i>	<b>0.329</b>	<b>0.308</b>	<b>15.216</b>	<b>0.000</b>	<b>223.715</b>	<b>1.971</b>

- **Selected SLR models**

The  $F$ -statistics value of the SLR models developed using project scope and quantity of work item predictor variables found to be statistically significant but the Durbin Watson statistics for the assumption of independent residuals of project scope does not meet the required value. Equations 27, 28, and 29 below show the simple linear regression structural framework duration prediction models that satisfied the regression assumptions listed previously.

$$\text{SLR1} \quad D_{Total} = 405.653 + 0.001Q_{bar} \quad (27)$$

$$\text{SLR2} \quad D_{Super} = 316.439 + 0.039Q_{exc} \quad (28)$$

$$\text{SLR3} \quad D_{Sub} = 78.073 + 0.001Q_{bar} \quad (29)$$

#### 4.5.1.1 Interpretation of selected SLR models

The coefficient of determination ( $R^2$ ) used to compare the performance of SLR models that satisfied assumptions of linear regression. Regression results indicated that quantity of work item variable  $Q_{bar}$  explained 34.9% of significant variation in duration of  $D_{Total}$  and 32.9% of  $D_{Sub}$ . While 25.4% significant variance in  $D_{Super}$  was explained by the single variable  $Q_{exc}$ . The amount of variability in structural frameworks construction duration accounted by selected SLR models can be considered as low. This indicates that significant portion of the variability is explained by other predictor variables.

Interpretation of coefficients regression models should be practical and convincing specially the constant. There might be instances when it becomes impractical for the predictor variable to have a zero value.

The coefficients of the  $Q_{bar}$  in SLR models indicated in equations 27 and 29 of  $D_{Total}$  and  $D_{Sub}$  indicated that an increase in quantity of bar by 1000 kilograms units increase the total and substructure duration by a one day. It takes nearly 405 days in total duration when there is no quantity of bar work, this might be the duration consumed by other structural framework constructions. While this value is 78 days for substructure duration. An increase in quantity of excavation by 1000 cubic meter units lead to an increase in 39 days of superstructure works. The interpretation of the constant term in  $D_{Super}$  equation would be impractical as there will be no project with zero quantities of excavation.

#### 4.5.1.2 Improved SLR Models

Mathematical transformation helps to increase the amount of variability explained by an independent variable on the dependent variable of interest. Further also avoid violations of linearity assumptions. But the interpretation and comparison of developed models become impractical. Table 4-22 below shows comparison of different types of SLR models based on transformation of the dependent variable  $D_{Total}$ .

**Table 4-22 Comparison of SLR models for  $D_{Total}$**

Dependent Variable	Independent Variable	R <sup>2</sup>	R <sup>2</sup> adjusted	F-value	Significance	Standard error of estimate	Durbin-Watson
<i>INV_D_Total</i>	<i>T_Floors</i>	0.123	0.095	4.366	0.045	0.00188	1.041
<i>Log_D_Total</i>	<i>T_Height</i>	0.143	0.115	5.161	0.030	0.32372	1.070
<i>Log_D_Total</i>	<i>T_GFA</i>	0.231	0.207	9.328	0.005	0.30653	1.151
<i>SQRT_D_Total</i>	<i>Ty_area</i>	0.267	0.244	11.300	0.002	8.34928	1.162
<i>Log_D_Total</i>	<i>Q_excV</i>	0.308	0.286	13.807	0.001	0.29081	1.639
<i>INV_D_Total</i>	<i>Q_form</i>	0.291	0.268	12.726	0.001	0.00169	1.099
<b><i>D_Total</i></b>	<b><i>Q_bar</i></b>	<b>0.349</b>	<b>0.328</b>	<b>16.621</b>	<b>0.000</b>	<b>432.4057</b>	<b>1.933</b>
<b><i>SQRT_D_Total</i></b>	<b><i>Q_bar</i></b>	<b>0.349</b>	<b>0.328</b>	<b>16.653</b>	<b>0.000</b>	<b>7.86637</b>	<b>1.775</b>
<i>Log_D_Total</i>	<i>Q_conc</i>	0.132	0.104	4.719	0.038	0.32571	1.108

Similar mathematical transformations conducted for superstructure and substructure durations and results of selected models presented in Table 4-23 and Table 4-24 below. A full list of 37 SLR models developed for the three dependent variables together with both statistical performance and corresponding t-statistics test for the coefficients presented in the appendixes section.

**Table 4-23 Comparison of SLR models for  $D_{Super}$** 

Dependent Variable	Independent Variable	R <sup>2</sup>	R <sup>2</sup> adjusted	F-value	Significance	Standard error of estimate	Durbin-Watson
<i>Log_D_Super</i>	<i>T_Floors</i>	0.169	0.142	6.302	0.017	0.30569	1.190
<i>Log_D_Super</i>	<i>B_Floors</i>	0.129	0.101	4.596	0.040	0.31293	1.445
<i>INV_D_Super</i>	<i>A_Floors</i>	0.130	0.102	4.638	0.039	0.00265	1.116
<i>Log_D_Super</i>	<i>T_Height</i>	0.224	0.199	8.942	0.005	0.29542	1.032
<i>D_Super</i>	<i>T_GFA</i>	0.289	0.266	12.579	0.001	278.2064	1.328
<i>D_Super</i>	<i>Ty_area</i>	0.285	0.261	12.331	0.001	279.0011	1.278
<b><i>Log_D_Super</i></b>	<b><i>Q_excv</i></b>	<b>0.286</b>	<b>0.263</b>	<b>12.444</b>	<b>0.001</b>	<b>0.28326</b>	<b>1.541</b>
<i>INV_D_Super</i>	<i>Q_form</i>	0.169	0.143	6.322	0.017	0.00259	1.345
<i>Log_D_Super</i>	<i>Q_bar</i>	0.257	0.233	10.734	0.003	0.28901	1.490
<i>D_Super</i>	<i>Q_conc</i>	0.123	0.095	4.342	0.046	308.9277	1.265

**Table 4-24 Comparison of SLR models for  $D_{Sub}$** 

Dependent Variable	Independent Variable	R <sup>2</sup>	R <sup>2</sup> adjusted	F-value	Significance	Standard error of estimate	Durbin-Watson
<i>INV_D_Sub</i>	<i>B_Floors</i>	0.238	0.213	9.685	0.004	0.00856	1.109
<i>Log_D_Sub</i>	<i>Ty_area</i>	0.513	0.497	32.633	0.000	0.00685	1.312
<i>INV_D_Sub</i>	<i>INV_Q_excv</i>	0.420	0.401	22.451	0.000	0.00747	1.446
<i>INV_D_Sub</i>	<i>INV_Q_form</i>	0.541	0.526	36.553	0.000	0.00665	1.186
<i>INV_D_Sub</i>	<i>INV_Q_bar</i>	0.162	0.135	6.001	0.020	0.42899	1.127
<b><i>D_Sub</i></b>	<b><i>Q_bar</i></b>	<b>0.329</b>	<b>0.308</b>	<b>15.216</b>	<b>0.000</b>	<b>223.715</b>	<b>1.971</b>

#### 4.5.2 Multiple Linear Regression (MLR) Models

In order to increase the amount of variability explained by SLR model's multiple linear regression analysis was conducted on SPSS. As discussed in the literature review chapter MLR consists of different methods of selection of predictor variables. Backward elimination, forward selection and stepwise method on SPSS were used. Table 4-25 shows the predictive performance of MLR models developed.

**Table 4-25 Predictive Performance of Multiple Linear Regression (MLR) Models**

Method	Dependent Variable	R <sup>2</sup>	R <sup>2</sup> adjusted	F-value	Sig.	Standard error of estimate	Durbin- Watson
Backward elimination	<i>D_Total</i>	0.546	0.481	8.416	0.000	379.9943	1.873
	<i>D_Super</i>	0.505	0.435	7.153	0.000	244.0942	1.442
	<i>D_Sub</i>	0.394	0.354	9.757	0.001	216.136	1.971
Forward selection	<i>D_Total</i>	0.472	0.436	13.391	0.000	395.9902	1.971
	<i>D_Super</i>	0.397	0.357	9.896	0.001	260.2718	1.439
	<i>D_Sub</i>	0.329	0.308	15.216	0.000	223.716	1.971
Stepwise	<i>D_Total</i>	0.472	0.436	13.391	0.000	395.990	1.971
	<i>D_Super</i>	0.397	0.357	9.896	0.001	260.272	1.439
	<i>D_Sub</i>	0.329	0.308	15.216	0.000	223.716	1.971

SPSS software uses different criteria for inclusion and exclusion of a predictive variable into a model and present a statistically significant model with improved performance. Table 4-26 provides the statistical data of the regression coefficients of the MLR models developed based on the three techniques.

- **Selected MLR models**

Output of forward selection and stepwise method are resulted in similar models. While the Durbin-Watson statistics for independent residuals is out of the recommended range for MLR models of *D\_Super* for all the three methods. Although The MLR model for *D\_Total* developed using backward elimination have higher adjusted coefficient of determination but it was discarded due to the possible multicollinearity which is possible linear association among predictor variables *A\_Floors* and *T\_Height*. Finally, two MLR models for *D\_Total* and *D\_Sub* are selected and presented in equations 30 and 31.

$$\text{MLR1} \quad D\_Total = 243.381 + 0.001Q\_bar + 0.047Q\_excv \quad (30)$$

$$\text{MLR2} \quad D\_Sub = 17.917 + 0.001Q\_bar + 0.017Q\_excv \quad (31)$$

**Table 4-26 Statistical data of the regression coefficients of MLR models**

Method	Dependent variable	Unstandardized Coefficients		Standardised Coefficients			Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Tolerance	VIF
Backward elimination method	(Constant)	149.251	215.525		0.692	0.494		
	<i>D_Total</i> <i>A_Floors</i>	-107.51	56.665	-0.664	-1.9	0.068	0.132	7.55
	<i>T_Height</i>	30.231	14.191	0.765	2.13	0.042	0.126	7.944
	<i>Q_excV</i>	0.04	0.018	0.319	2.244	0.033	0.801	1.248
	<i>Q_bar</i>	0.001	0	0.388	2.765	0.010	0.825	1.212
	(Constant)	28.508	138.445		0.206	0.838		
	<i>D_Super</i> <i>A_Floors</i>	-67.333	36.399	-0.676	-1.850	0.075	0.132	7.55
	<i>T_Height</i>	23.192	9.116	0.953	2.544	0.017	0.126	7.944
	<i>Q_excV</i>	0.021	0.011	0.273	1.842	0.076	0.801	1.248
	<i>Q_bar</i>	0.000	0.000	0.261	1.781	0.086	0.825	1.212
	<i>D_Sub</i> (Constant)	17.917	69.847		0.257	0.799		
	<i>Q_excV</i>	0.017	0.010	0.272	1.792	0.083	0.874	1.144
<i>Q_bar</i>	0.001	0.000	0.477	3.141	0.004	0.874	1.144	
Forward selection method	(Constant)	243.381	127.970		1.902	0.067		
	<i>D_Total</i> <i>Q_bar</i>	0.001	0.000	0.458	3.228	0.003	0.874	1.144
	<i>Q_excV</i>	0.047	0.018	0.374	2.639	0.013	0.874	1.144
	(Constant)	203.56	85.177		2.39	0.023		
	<i>D_Super</i> <i>T_GFA</i>	0.016	0.006	0.407	2.668	0.012	0.865	1.157
	<i>Q_excV</i>	0.027	0.012	0.355	2.328	0.027	0.865	1.157
	(Constant)	78.073	63.403		1.231	0.227		
	<i>D_Sub</i> <i>Q_bar</i>	0.001	0.000	0.574	3.901	0.000	1	1
	(Constant)	243.381	127.97		1.902	0.067		
	<i>D_Total</i> <i>Q_bar</i>	0.00100	0.000	0.458	3.228	0.003	0.874	1.144
<i>Q_excV</i>	0.047	0.018	0.374	2.639	0.013	0.874	1.144	
Stepwise method	(Constant)	203.560	85.177		2.390	0.023		
	<i>D_Super</i> <i>T_GFA</i>	0.016	0.006	0.407	2.668	0.012	0.865	1.157
	<i>Q_excV</i>	0.027	0.012	0.355	2.328	0.027	0.865	1.157
	(Constant)	78.073	63.403		1.231	0.227		
	<i>D_Sub</i> <i>Q_bar</i>	0.001	0.000	0.574	3.901	0.000	1	1

17 more MLR models are developed using combinations of a maximum of three predictor variables and statistical results are presented in the appendixes section. Equations 32, 33, and 34 are selected MLR models that satisfy the regression assumptions.

$$\text{MLR3 } D\_Total = 143.52 + 0.001Q\_bar + 0.043Q\_excV + 0.217Ty\_area \quad (32)$$

$$\text{MLR4 } D\_Super = 165.956 + 0.023Q\_excV + 0.00029Q\_bar + 0.013T\_GFA \quad (33)$$

$$\text{MLR5 } D\_Sub = 168.059 + 0.02Q\_excV + 0.001Q\_bar - 15.183T\_Floors \quad (34)$$

**4.5.2.1 Predictive models validation and selection**

A total of 37 SLR and 17 MLR models were developed, after conducting tests for possible violations of basic assumptions for each respective model eight of them were selected for final comparison. The eight predictive models validated using the validation data set. Comparison of the models were conducted using the adjusted  $R^2$ , standard error of estimate and the root mean square error (RMSE). Table 4-27 below presents the comparison of the models and selected final predictive models are highlighted in bold.

**Table 4-27 Predictive models comparison and validation statistics**

Dependent variable	Model	Model data set			Validation data set
		R <sup>2</sup> adjusted	Standard error of estimate	RMSE	RMSE
<i>D_Total</i>	SLR1	0.328	432.405	429.239	508.131
	MLR1	0.436	395.990	378.193	465.542
	<b>MLR3</b>	<b>0.452</b>	<b>390.371</b>	<b>381.921</b>	<b>451.162</b>
<i>D_Super</i>	SLR2	0.230	284.805	276.054	322.716
	<b>MLR4</b>	<b>0.379</b>	<b>255.769</b>	<b>239.778</b>	<b>286.494</b>
<i>D_Sub</i>	SLR3	0.308	223.715	269.195	339.026
	MLR2	0.354	216.136	302.174	337.934
	<b>MLR5</b>	<b>0.369</b>	<b>213.569</b>	<b>276.663</b>	<b>344.125</b>

45.2% variance in total structural frameworks duration explained using the predictive model MLR3. While 37.9% variance in superstructure and 36.9% in substructure duration explained by MLR4 and MLR5 respectively. The two quantity of work item variables (i.e.  $Q_{bar}$ , and  $Q_{exc}$ ) appeared in all the three predictive MLR models. while project scope variables  $T_{GFA}$ ,  $T_{y\_area}$ , and  $T_{Floors}$  explained significant portion of the variance of the dependent variables.

Table 4-28 shows the values of the predicted duration using the selected models versus the observed structural works duration.

**Table 4-28 Predicted durations of validation data set projects**

ID	Observed durations			Predicted durations		
	<i>D_Total</i>	<i>D_Super</i>	<i>D_Sub</i>	MLR3	MLR4	MLR5
				<i>D_Total</i>	<i>D_Super</i>	<i>D_Sub</i>
Apat-036	1,200	1,055	145	1,247	646	693
Apat-037	725	650	75	780	475	331
Apat-039	1,300	600	700	1,420	645	821
Apat-057	980	750	230	1,000	813	155
Apat-074	182	110	60	447	380	58
Apat-107	470	350	120	751	482	203
Apat-167	340	240	100	1,060	515	679
Apat-168	1,601	1,011	590	632	475	123

### 4.5.3 Best-fit Linear Regression Models

The main objective of this study is to develop a duration prediction model, that can be utilized for future estimation. Mathematical models need to be capable of explaining a larger portion of the variance in the dependent variable to achieve this intended purpose. The simple and multiple linear regression techniques presented in the previous two subsections however, resulted in models with low coefficient of determination. The amount of variance in the dependent structural works duration explained by the predictor variables is low. Therefore, additional mathematical computations carried out to develop a linear regression model that fits the data best.

- **Procedure**

To develop a best-fit model, the model building data set which contains 33 apartment projects is divided into two groups. The total gross floor area of buildings was used to split the data into two groups as; small-size and large-size buildings.

Small-size apartment	Gross floor area $\leq 8,000\text{m}^2$
Large-size apartment	Gross floor area $> 8,000\text{m}^2$

Table 4-29 below shows the data samples used for the best-fit model development. After splitting the data, a correlation analysis was conducted to identify the significant linear relation that exist between the dependent variables and predictor variables. Then, a simple scatter plot was plotted showing the linear relation that exist and to identify outlier data points. Detected outlier points were then removed from the model building data set. Then,

simple and multiple linear regression was carried out to develop a best fit model with an improved coefficient of determination.

**Table 4-29 Data splitting based on gross floor area and outliers removed**

<b>Small-size Apartment samples</b>		<b>Large-size Apartment Samples</b>	
Samples ID	Outliers removed based on $Q_{bar}$	Samples ID	Outliers removed based on $Ty_{area}$
Apat-286	Apat-288	Apat-166	Apat-196
Apat-056		Apat-287	Apat-011
Apat-289		Apat-073	Apat-012
Apat-144		Apat-201	Apat-106
Apat-143		Apat-196	
Apat-288		Apat-197	
Apat-203		Apat-011	
Apat-236		Apat-238	
Apat-072		Apat-061	
Apat-264		Apat-013	
Apat-262		Apat-222	
Apat-263		Apat-237	
Apat-265		Apat-071	
Apat-038		Apat-012	
Apat-202		Apat-221	
Apat-108		Apat-169	
		Apat-106	

---

Statistical data of the samples used for this best-fit model development is presented in the appendices section of this thesis.

#### **4.5.3.1 Best-fit models for small-sized apartment projects**

Small-sized apartment projects as defined here are projects with total gross floor area less than or equal to 8,000 square meters. Linear models with a capability of explaining up to 80 percent variance were developed using both SLR and MLR techniques. Selected best-fit MLR models are presented below. Further, table showing the statistical performance of linear models developed are presented in Table 4-30.

$$D_{Total\_s} = 508.482 - 20.176Q_{bar} + 0.065T_{GFA} + 0.002T_{Height} \quad (32)$$

$$D_{Super\_s} = 2003.84 + 0.001Q_{bar} - 568.11Ty_{height} - 0.014Q_{conc} \quad (32)$$

$$D_{Sub\_s} = -7.588 + 0.001Q_{bar} + 0.014Q_{form} - 0.017Q_{exc} \quad (32)$$

**Table 4-30 Best-fit models statistical performance for small-size apartments**

Dependent Variable	Model type	Independent Variable	R <sup>2</sup>	R <sup>2</sup> adjusted	F-value	Sig.	Standard error of estimate
<i>D_Total_s</i>	MLR	<i>T_Height</i>	0.741	0.67	10.484	0.001	260.7243
		<i>T_GFA</i>					
		<i>Q_bar</i>					
<i>D_Super_s</i>	SLR	<i>Q_bar</i>	0.691	0.667	29.017	0.000	262.0672
		<i>Q_bar</i>					
		<i>Ty_height</i>					
<i>D_Sub_s</i>	MLR	<i>Q_bar</i>	0.808	0.756	15.42	0.000	117.8482
		<i>Q_form</i>					
		<i>Q_excw</i>					
	SLR	<i>Q_bar</i>	0.654	0.627	24.549	0.000	145.5299

#### 4.5.3.2 Best-fit models for large-size apartment projects

Large-sized apartment projects as defined here are projects with total gross floor area greater than 8,000 square meters. Linear models with a capability of explaining up to 85 percent variance were developed using both SLR and MLR techniques. Selected best-fit MLR models are presented below. Further, table showing the statistical performance of linear models developed are presented in Table 4-31.

$$D\_Total\_L = -187.48 + 4.943T\_Height + 1.197Ty\_area - 0.029Q\_form \quad (32)$$

$$D\_Super\_L = 455.168 + 56.604T\_Height - 201.23T\_Floors + 0.034Q\_conc \quad (32)$$

$$D\_Sub\_L = -139.94 + 0.413Ty\_area - 0.011Q\_form + 0.025Q\_excw \quad (32)$$

**Table 4-31 Best-fit models statistical performance for large-size apartments**

Dependent Variable	Model type	Independent Variable	R <sup>2</sup>	R <sup>2</sup> adjusted	F-value	Sig.	Standard error of estimate
<i>D_Total_L</i>	MLR	<i>T_Height</i>	0.850	0.80	17.022	0.000	232.998
		<i>Ty_area</i>					
		<i>Q_form</i>					
<i>D_Super_L</i>	SLR	<i>Ty_area</i>	0.758	0.736	34.404	0.000	267.9916
		<i>T_Height</i>					
		<i>T_Floors</i>					
<i>D_Sub_L</i>	MLR	<i>Q_form</i>	0.756	0.674	9.28	0.004	149.9595
		<i>Q_excw</i>					
		<i>Ty_area</i>					
	SLR		0.594	0.557	16.071	0.002	174.9346

#### 4.5.3.3 Validation of best-fit models

Cross validation of developed models was conducted using validation samples which are not used for model development purpose. From the randomly selected eight samples only one sample is small-size while the other are large-size. Data of the observed and predicted structural works durations are presented in Table 4-32.

**Table 4-32 Validation of best-fit models**

Apartment group	Apartment ID	Observed durations			Predicted durations			
		<i>D_Total</i>	<i>D_Super</i>	<i>D_Sub</i>	<i>D_Total</i>	<i>D_Super</i>	<i>D_Sub</i>	
Small-size	Apat-037	725	650	75	410	547	217	
					RMSE	315	103	142
Large-size	Apat-167	340	240	100	337	634	36	
Large-size	Apat-074	182	110	60	849	1215	185	
Large-size	Apat-039	1300	600	700	1429	823	474	
Large-size	Apat-168	1601	1011	590	696	224	168	
Large-size	Apat-107	470	350	120	1500	716	441	
Large-size	Apat-036	1200	1055	145	1672	311	509	
Large-size	Apat-057	980	750	230	1573	514	575	
Large-size	Apat-057	980	750	230	1573	514	575	
					RMSE	683	638	321

## 4.6 Comparison of Different Construction Methods

### 4.6.1 Non-parametric tests for comparing difference between groups

The final research question addressed the comparison of groups of construction method variables assumed to influence the duration of structural frameworks of buildings. Mann-Whitney U Test and Kruskal-Wallis Test which are the nonparametric equivalents of independent samples t-test and one-way ANOVA between groups used.

#### 4.6.1.1 Description of data set

The structural framework construction duration survey used to gather sample data of 41 previously constructed residential apartment projects in Addis Ababa. One of the samples was identified as extreme outlier from box plot the dependent variables with selected categorical variables and removed leaving 40 samples for comparing means of different groups. Table 4-33 below shows the frequencies of the construction method factor groups.

**Table 4-33 Frequency table of construction method factor groups**

Construction method factor	Symbol	Group	Frequency	Percent	Valid Percent	Cumulative Percent
Type of foundation	<i>Found_ty</i>	Isolated/combined footing	33	82.5	82.5	82.5
		Raft/Mat foundation	7	17.5	17.5	100
		Total	40	100	100	
		<hr/>				
Concrete mixing method	<i>Conc_method</i>	On site mixing	29	72.5	72.5	72.5
		Ready mix delivered with truck	11	27.5	27.5	100
		Total	40	100	100	
		<hr/>				
Use of tower crane	<i>Tower_C</i>	Tower crane used	3	7.5	7.5	7.5
		Tower crane not used	37	92.5	92.5	100
		Total	40	100	100	
		<hr/>				
Shoring & scaffoldig sistem	<i>Form_sys</i>	Metal	8	20	20	20
		Combination	32	80	80	100
		Total	40	100	100	
		<hr/>				
Use of concrete pump	<i>Conc_pump</i>	Pump used	10	25	25	25
		Pump not used	30	75	75	100
		Total	40	100	100	
		<hr/>				
Type of slab	<i>Slab_ty</i>	Conventional 1 way or 2 way slab	23	57.5	57.5	57.5
		Flat slab	10	25	25	82.5
		Ribbed slab	7	17.5	17.5	100
		Total	40	100	100	
		<hr/>				

#### **4.6.1.2 Statistical procedure for analysis**

Non-parametric equivalent of Independent samples t-test (i.e. Mann-Whitney U Test) and one-way ANOVA (i.e. Kruskal-Wallis Test) was used to compare groups of construction method factors. The necessity of Non-parametric statistic arises due to the unbalanced sample size in each groups which lead to the violation of homogeneity of variance between groups. Further also the unbalanced sample sizes create the distribution of the dependent variable to deviate from the required normal distribution.

Mann-Whitney U Test was used to test for differences between two independent groups of construction method related factors on the continuous measures of structural framework construction duration as measured with total, superstructure and substructure duration. While Kruskal-Wallis test used to compare deference between three groups.

#### **4.6.1.3 Results of Independent Mann-Whitney U Test**

The construction method factors: Foundation type, concrete mixing method, use of tower crane, shoring and scaffolding system, and use of concrete pump were analyzed for differences between two independent groups in each factor on duration of structural frameworks construction. Mann-Whitney U Test converts the durations in each group into ranks and compares if the ranks for the two groups differ significantly. The mean and sum of the ranks together with the test statistics are presented for the five construction method factors.

- **Foundation type**

A Mann-Whitney U test revealed no significant difference in the construction duration of total, super structure and substructure works of buildings with isolated/combined footing and raft/mat foundation. The Z value read from the statics table is -1.228, -0.677, and -1.835 for  $D_{Total}$ ,  $D_{Super}$ , and  $D_{Sub}$  respectively. The respective p values are 0.219, 0.499, and 0.066 for  $D_{Total}$ ,  $D_{Super}$ , and  $D_{Sub}$  respectively. All p values are greater than the significant value 0.05 so the results are not significant. Table 4-34 below shows the rank and the test statistics.

**Table 4-34 Mann-Whitney U Test Statistics for Foundation type**

Foundation Type (Found_ty)		N	Mean Rank	Sum of Ranks
<i>D_Total</i>	Isolated/combined footing	33	21.55	711
	Raft/Mat foundation	7	15.57	109
	Total	40		
<i>D_Super</i>	Isolated/combined footing	33	21.08	695.5
	Raft/Mat foundation	7	17.79	124.5
	Total	40		
<i>D_Sub</i>	Isolated/combined footing	33	22.06	728
	Raft/Mat foundation	7	13.14	92
	Total	40		

Test Statistics

	<i>D_Total</i>	<i>D_Super</i>	<i>D_Sub</i>
Mann-Whitney U	81.000	96.500	64.000
Wilcoxon W	109.000	124.500	92.000
Z	-1.228	-.677	-1.835
Asymp. Sig. (2-tailed)	.219	.499	.066

- **Concrete mixing method**

The effect of particular group of concrete mixing method on the dependent variable construction duration was assessed using Mann-Whitney U test. No significant difference was observed on construction duration of apartment projects for concrete on site mixing and ready mix delivered with truck. Output of the test statistics are presented in the appendixes section of the thesis.

- **Use of tower crane**

Three apartment projects out of the total 40 used tower crane for transportation of construction materials but the Mann-Whitney test conducted indicated that there was no significant difference on the construction duration from the rest projects which don't use. The statistical test results accessible in the appendixes section.

- **Shoring and scaffolding system**

There groups of shoring and scaffolding system (i.e. eucalyptus wood, metal, and combination) were created in the questionnaire to compare possible variances in

construction duration between groups. Eight used metal only system while 32 used combination where the eucalyptus system was discarded due to no sample and comparison done with only two groups. No significant difference was observed on durations of projects with metal and combination system statistical results presented in the appendixes section.

- **Use of concrete pump**

A Mann-Whitney U test indicated that there was no significant difference on structural framework durations between apartment projects that used concrete pump and those that don't. Statistical tables are available in the appendixes section.

**4.6.1.4 Result of Kruskal-Wallis test**

Kruskal-Wallis test is the non-parametric equivalent of one-way ANOVA between groups. Here the number of groups to compare are three and more. The type slab construction method factor with conventional 1-way or 2-way, flat slab, and ribbed slab was analyzed for significant difference in construction duration of structural framework of building. Durations are converted to rank and the mean rank in each group compared to identify significant difference. Table 4-35 below shows the ranks of the construction durations of samples in the three slab type groups. Mean rank of ribbed slab is greater than both conventional 1-way 2-way and flat slab but the chi-square result revealed this difference was not significant.

**Table 4-35 Mean Rank values of durations in slab type categories**

	Slab_ty	N	Mean Rank
<i>D_Total</i>	Conventional 1 way or 2 way slab	23	21.57
	Flat slab	10	16.85
	Ribbed slab	7	22.21
	Total	40	
<i>D_Super</i>	Conventional 1 way or 2 way slab	23	20.76
	Flat slab	10	18.30
	Ribbed slab	7	22.79
	Total	40	
<i>D_Sub</i>	Conventional 1 way or 2 way slab	23	21.17
	Flat slab	10	16.80
	Ribbed slab	7	23.57
	Total	40	

The Kruskal-Wallis test statistic presented below in Table 4-36 revealed that there is no significant difference in structural frame work durations (i.e.  $D_{Total}$ ,  $D_{Super}$ ,  $D_{Sub}$ ) across the three different slab types.

**Table 4-36** Kruskal-Wallis test statistic for slab types

	$D_{Total}$	$D_{Super}$	$D_{Sub}$
Chi-Square	1.317	.634	1.565
df	2	2	2
Asymp.	.518	.728	.457
Sig.			

#### 4.6.1.5 Discussions on the results of RQ-3

The third question in this research used non-parametric tests on the selected Construction method independent variables and revealed that there was no statistically significant difference in construction duration across groups. This result however, gives a helpful hint for future research that focus on durations of buildings in the same setting as the one used in this research.

## CHAPTER 5 CONCLUSIONS AND RECOMMENDATIONS

### 5.1 Conclusions

This study provides structural frameworks construction duration prediction models that can be used by contractors, clients, and consultants that explained significant portions of variance. Significant linear relationship was observed between quantity of work item variables which are new additions to the former project scope variables. The use of quantity of work items in duration estimation models eliminates the requirement of adjustment of models developed using cost which is the common project scope variable used previously.

This research performed; correlation analysis using Pearson's correlation coefficient, hypothesis testing to determine existence of linear relation, simple and multiple linear regression analysis to develop prediction models and comparison of groups of construction method.

Two research questions together with two hypotheses helped to achieve the study main objective of determining the interrelationship that exist between dependent structural works duration and predictor variables.

The first research question helped to identify the kind and degree of relationship that exist among the dependent structural framework construction duration variables (i.e. total, substructure & super structure construction durations) and each the selected project scope parameters (i.e. Gross and typical floor area, total number of floors above and below ground, total and typical building height.). A statistically significant moderate positive correlation was obtained for typical floor area with the three duration variables. While similar relation observed between total gross floor area and superstructure and total structural works duration. Total number of floors and total building height showed significant moderate positive linear relation with superstructure duration.

This research question was converted to testable hypothesis stating; there is a significant linear relation between structural works duration and selected project scope factors (i.e. Gross and typical floor area, total number of floors above and below ground, total and typical building height.) The null hypothesis for no significant linear relation was rejected

for the three dependent duration variables based on the t-statistics of the slope coefficients of simple linear regression.

The second research question intended to identify the kind and degree of association among the dependent structural framework construction duration variables (i.e. total, substructure & super structure construction durations) and each of independent selected quantities of work items (i.e. quantity of concrete, excavation, formwork and rebar). The Pearson's correlation coefficient revealed that, the four quantity of work item variables showed statistically significant positive linear relation with all duration variables except for quantity of concrete where significant relation was observed only with superstructure duration.

Similar to the first research question this one also converted to testable hypothesis stating; there is a significant relation between structural works duration and selected quantity of work elements (i.e. volume of concrete, excavation, quantity of formwork and rebar). The null hypothesis of no significant linear relation was rejected and the alternative hypothesis accepted using the p-values of the t-statistics of the coefficients of the simple linear regression.

After addressing the first two sets of questions and hypotheses structural framework duration estimation models were developed using simple and multiple linear regression. 37 simple linear regression and 17 multiple linear regression model are developed and corresponding statistical results provided in the appendixes section.

The amount of variance as measured by the coefficient of determination was observed to be low and another set of linear regression was carried out to develop best-fit model. by splitting the data set into small-size and large-size apartments based on total gross floor area and removing outlier samples linear models with higher coefficients of determination obtained.

A surprising result was obtained from non-parametric tests conducted to compare different groups of construction method suggesting that there was no significant difference in duration among groups.

## **5.2 Recommendations**

### **5.2.1 Recommendations for project stakeholders**

Developed models should be used only for data that are within the range of those that were used for model building purpose.

The detailed estimation of quantity of work items gives a head start for prediction of duration where next productivity can be calculated for resource deployment.

Unavailability of record of project history was observed in all real-estate developing companies. Companies should develop a simple project history data base.

Specialization of structural framework contractors would enhance the productivity of the construction industry and management of projects.

### **5.2.2 Future research areas**

This study only conceptualizes on a handful of structural framework construction duration affecting factors specifically quantitative ones, fellows are encouraged to investigate other factors. Especially incorporation of qualitative factors in predictive models would add a value to theory and practical use.

Similar studies can be carried out on villa and commercial buildings in a separate setting which were not done here due to small sample sizes.

## REFERENCES

- AbouRizk, S., 2010. Role of Simulation in Construction Engineering and Management. *J. Constr. Eng. Manage.* , Volume 136, pp. 1140-1153.
- AbouRizk, S., Halpin, D., Mohamed, Y. & Hermann, U., 2011. Research in Modeling and Simulation for Improving Construction Engineering Operations. *J. Constr. Eng. Manage.*, 137(10), pp. 843-852.
- Ahmadu, H. A., Ibrahim, Y. M., Ibrahim, A. D. & Abdullahi, M., 2015. Modelling building construction durations. *Journal of Financial Management of Property and Construction*, Volume 20, pp. 65-84.
- Ahuja, H. N. & Nandakumar, V., 1985 . Simulation Model To Forecast Project Completion Time. *Journal of Construction Engineering and Management*, Volume 111, pp. 325-342.
- Ameyaw, C., Mensah, S. & Arthur, Y. D., 2012. *Applicability of Brownlow's Time – Cost Model On Building Projects in Ghana*. Abuja, Nigeria,, West Africa Built Environment Research (WABER) Conference, pp. 881-888.
- Andinet, Behailu & Derbachew, 2006. *Time Cost Relationship: Educational Building Projects in Ethiopia.* , , AAU, Ethiopia: Final year project.
- Assefa, A., 2008. *Time – Cost Relationships for Public Road Construction Projects in Ethiopia*. Addis Ababa , Ethiopia: MSc. Thesis, Addis Ababa University.
- Ayyub, B. M. & Haldar, A., 1984. Project Scheduling Using Fuzzy Set Concepts. *Journal of Construction Engineering and Management*, Volume 110, pp. 189-204.
- Baldwin, A. & Bordoli, D., 2014. *A Handbook for Construction Planning and Scheduling*. 1st ed. The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK: John Wiley & Sons, Ltd.
- Bang, J.-D. et al., 2011. Time Shortening of Structural Framework of High-rise Apartment Housing for the Urgent Project Area: Focused on the Cases of Insufficient Time or Delayed Projects. *LHI Journal*, 2(4), pp. 355-365.

Barton, B. & Peat, J., 2014 . *Medical statistics : a guide to SPSS, data analysis, and critical appraisal*. 2nd ed. The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK: John Wiley & Sons Ltd.

Barton, B. & Peat, J., 2014. *Medical Statistics: A Guide to SPSS, Data Analysis and Critical Appraisal*. 2nd ed. The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK: John Wiley & Sons Ltd..

Bayram, S., 2016. Duration Prediction Models for Construction Projects:In Terms of Cost or Physical Characteristics?. *KSCE Journal of Civil Engineering*, pp. 1-12.

Bennett, J. & Ormerod, R. N. ..., 1984. Simulation applied to construction projects. *Construction Management and Economics*, Volume 2, pp. 225-263.

Bhokha, S. & Ogunlana, S. O., 1999 . Application of artificial neural network to forecast construction duration of buildings at the predesign stage. *Engineering, Construction and Architectural Management* , 6(2), pp. 133-144 .

Birhane, A., 2015. *Relationships of construction time and cost for railway projects in Ethiopia*. Addis Ababa, Ethiopia: MSc. Thesis, Addis Ababa University .

Bokor, O., Florez, L., Osborne, A. & Gledson, B. J., 2019. Overview of construction simulation approaches to model construction processes. *Organization, Technology and Management in Construction* , Volume 11, p. 1853–1861.

Botchkarev, A., 2019. A new typology design of performance metrics to measure errors in machine learning regression algorithms. *Interdisciplinary Journal of Information, Knowledge, and Management*, Volume 14, pp. 45-79.

Boussabaine, A. H., 2001. Neurofuzzy modelling of construction projects' duration I: principles. *Engineering, Construction and Architectural Management* , 8(2), p. 104–113.

Bromilow, F., 1969. Contract time performance expectations and the reality. *Building Forum*, 1(3), pp. 70-80.

Bromilow, F., Hinds, M. & Moody, N., 1980 . AIQS Survey of building contract time performance. *The Building Economist* , 19(2), pp. 79-82.

Brooks, G. P. & Barcikowski, R. S., 2012 . The PEAR Method for Sample Sizes in Multiple Linear Regression. *Multiple Linear Regression Viewpoints*, 38(2), pp. 1-16.

Chan, A. P., 1999. Modelling building durations in Hong Kong. *Construction Management and Economics*, Volume 17, pp. 189-196.

Chan, A. P., 2001. Time-Cost relationship of public sector projects in Malaysia. *International Journal of Project Management* , Volume 19, pp. 223-229.

Chan, A. P. C., Chan, D. W. M. & Yeung, J. F. Y., 2009. Overview of the Application of “Fuzzy Techniques” in Construction Management Research. *J. Constr. Eng. Manage.*, Volume 135, pp. 1241-1252.

Chan, D. W. & Kumaraswamy, M. M., 1995. A study of the factors affecting construction durations in Hong Kong. *Construction Management and Economics*, Volume 13, pp. 319-333.

Chan, D. W. & Kumaraswamy, M. M., 1999b. Modelling and predicting construction durations in Hong Kong public housing. *Construction Management and Economics*, Volume 17, pp. 351-362.

Chan, D. W. M. & Kumaraswamy, M. M., 1996a. A comparative study of time overruns in Hong Kong construction projects. *International Journal of Project Management*, 15(1), pp. 55-63.

Chan, D. W. M. & Kumaraswamy, M. M., 1996b. An Evaluation of Construction Time Performance in the Building Industry. *Building and Environment*, 31(6), pp. 569- 578.

Chatterjee, S. & Hadi, A. S., 2006. *Regression Analysis by Example*. 4th ed. New Jersey.: John Wiley & Sons, Inc., Hoboken,.

Chen, W. T. & Huang, Y.-H., 2006. Approximately predicting the cost and duration of school reconstruction projects in Taiwan. *Construction Management and Economics* , Volume 24, p. 1231–1239.

Choudhury, I., 2012. *A Study of the Factors of Construction Time for Educational Projects in Texas*. America, American Society for Engineering Education, pp. 1-9 .

Choudhury, I. & Rajan, S. S., 2004. *Time-Cost Relationship for Residential Construction in Texas*. Roanoke, Virginia., Proceedings of the 38th Annual Conference of the Associated Schools of Construction. .

Claudy, J. G., 1978 . Multiple Regression and Validity Estimation in One Sample. *Applied Psychological Measurement* , 2(4 ), pp. 595-607.

Dattalo, P., 2008 . *Determining Sample Size: Balancing Power, Precision, and Practicality*. 198 Madison Avenue, New York : Oxford University Press, Inc. .

Dawood, N., 1998. Estimating project and activity duration: a risk management approach using network analysis. *Construction Management and Economics*, Volume 16, pp. 41-48.

Dissanayaka, S. M. & Kumaraswamy, M. M., 1999 . Evaluation of factors affecting time and cost performance in Hong Kong building projects. *Engineering, Construction and Architectural Management*, 6( 3), pp. 287-298.

Dursun, O. & Stoy, C., 2011a. *An Evaluation of Construction Speed Performance for Building Projects in UK and Germany*. Amsterdam, The Netherlands, s.n.

Dursun, O. & Stoy, C., 2011b. Time–cost relationship of building projects: statistical adequacy of categorization with respect to project location. *Construction Management and Economics*, Volume 29, p. 97–106.

Edmonds, W. A. & Kennedy, T. D., 2017. *An Applied Guide to Research Designs Quantitative, Qualitative, and Mixed Methods*. 2nd ed. 2455 Teller Road Thousand Oaks, California : SAGE Publications, Inc .

Elhag, T. & Boussabaine, A., 1999. *Evaluation of Construction Cost and Time Attributes*. Liverpool L69 3BX, UK, Association of Researchers in Construction Management, pp. 473-80.

Ethiotelecom, 2016. *Federal Democratic Republic of Ethiopia (EFDRE) Official telephone directory*. 2016/17 ed. Addis Ababa, Ethiopia : Ethio Telecom.

Etikan, I., Musa, S. A. & Alkassim, R. S., 2016. Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics* , 5(1), pp. 1-4.

Fahrmeir, L., Kneib, T., Lang, S. & Marx, B., 2013. *Regression: Models, Methods and Applications*. Verlag Berlin Heidelberg : Springer-

Gab-Allah, A. A., Ibrahim, A. H. & Hagraas, O. A., 2015. Predicting the construction duration of building projects using artificial neural networks. *Int. J. Applied Management Science*, Volume 7, pp. 123-141.

Golizadeh, H., Sadeghifam, A. N., Aadal, H. & Majid, M. Z. A., 2015. Automated Tool for Predicting Duration of Construction Activities in Tropical Countries. *KSCE Journal of Civil Engineering*, 00(0), pp. 1-11 .

Guerrero, M. A., 2016. *Modelling Building Construction Speed by using Linear Regression Analysis, Artificial Neural Network, and n-Dimensional Finite elements*. s.l.:PHD Thesis, University of Alicante.

Guerrero, M. A., Villacampa, Y. & Montoyo, A., 2014. Modeling construction time in Spanish building projects. *International Journal of Project Management*, 32(2014), pp. 861-873.

Haykin, S., 1999. *Neural Networks: A comparative Foundation*. 2nd ed. Upper saddle River, New Jersey: Prentice Hall.

Hoffman, G. J., Jr., A. E. T., Webb, T. S. & Weir, J. D., 2007. Estimating Performance Time for Construction Projects. *Journal of Management in Engineering*, Volume 23, pp. 193-199.

Husin, A. E., 2019. Time Performance Upgrade by Critical Chain Project Management and BIM 4D Integration On Top Structural Work of a High Rise Building Construction Project. *ARPJ Journal of Engineering and Applied Sciences*, 14(17), pp. 3063-3072.

Ireland, V., 1983. *The role of managerial actions in the cost, time, and quality performance of high rise commercial building projects..* Sydney, NSW: Ph.D. Thesis, University of Sydney.

Jackson, S. L., 2009. *Research Methods and Statistics: A Critical Thinking Approach*. 3rd ed. 10 Davis Drive, Belmont, CA 94002-3098, USA: Wadsworth, Cengage Learning.

Jarkas, A. M., 2015. Predicting Contract Duration for Building Construction: Is Bromilow's Time-Cost Model a Panacea?. *Journal of Management in Engineering*, pp. 1-8.

Jin, R., Han, S., Hyun, C. & Cha, Y., 2016. Application of Case-Based Reasoning for Estimating Preliminary Duration of Building Projects. *Journal of Construction Engineering and Management*, Volume 142, pp. 1-8.

Kaka, A. & Prince, A., 1991. Relationship between value and duration of construction projects. *Construction Management and Economics*, Volume 9, pp. 383-400.

Khodakaram, V., Fenton, N. & Neil, M., 2007. Project Scheduling: Improved Approach to Incorporate Uncertainty Using Bayesian Networks. *Project Management Journal*, 38(2), pp. 39-49.

Khosrowshahi, F. & Kaka, A. P., 1996. Estimation of Project Total Cost and Duration for Housing Projects in the U.K. *Building and Environment*, pp. 373-383.

Kifle, S., 2015. *Construction Time Prediction for Public Office and Educational Building Projects in Addis Ababa, Ethiopia*. Mekelle University: A Thesis Submitted to The School of Civil Engineering.

Knofczynski, G. T. & Mundfrom, D., 2008. Sample Sizes When Using Multiple Linear Regression for Prediction. *Educational and Psychological Measurement*, 68 (3), pp. 431-442.

Kumar, R., 2011. *Research Methodology: a step-by-step guide for beginners*. 3rd ed. Mathura Road New Delhi: SAGE Publications Ltd.

Kumar, V. S. & Reddy, G., 2005. *Fuzzy Logic Approach to Forecast Project Duration in Construction Projects*. Construction Research Congress, ASCE.

Laptali, E., Bouchlaghem, N. & Wild, S., 1995. Construction of reinforced concrete multi-storey office buildings: a simulation model for time/cost calculations. *Building Research and Information*, 23 (4), pp. 227-233.

Lee, H.-s., Shin, J.-w., Park, M. & Ryu, H.-G., 2009. Probabilistic Duration Estimation Model for High-Rise Structural Work. *Journal of construction engineering and management*, Volume 135, pp. 1289-1298.

Le-Hoai, Lee, Y. D. & Cho, J. W., 2009 . *Construction of Time - Cost Model for Building Projects in Vietnam Long*. Busan, Proceedings of KICEM Conference.

Li, J., 2017 . Assessing the accuracy of predictive models for numerical data: Not r nor r2, why not? Then what?. *PLoS ONE* , 12(8 ), pp. 1-16.

Lim, J. H. et al., 2017 . Developing a Construction Duration Estimation Model to Ensure the Safety in Apartment Housing Construction Sites. *KSCE Journal of Civil Engineering* , Volume 0000, pp. 1-11.

Lin, M.-C., Tserng, H. P., Ho, S.-P. & Young, D.-L., 2011. Developing A Construction-Duration Model Based On A Historical Dataset for Building Project. *Journal Of Civil Engineering and Management* , Volume 17, p. 529–539.

Love, P. E. D., Tse, R. Y. C. & Edwards, D. J., 2005. Time–Cost Relationships in Australian Building Construction Projects. *JOURNAL OF CONSTRUCTION ENGINEERING AND MANAGEMENT*, pp. 187-194.

Mackova, D. et al., 2017. Construction-Duration Prediction Model for Residential Buildings in Slovak Republic Based on Computer Simulation. *International Journal of Applied Engineering Research* , pp. 3590-3599.

Maier, H. & Dandy, G., 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and applications. *Environmental Modeling & Software*, 15(1), pp. 101-124.

Maravas, A. & Pantouvakis, J., 2011. *A Process for the Estimation of the Duration of Activities in Fuzzy Project Scheduling 2011*. Vulnerability, Uncertainty, and Risk, ICVRAM and ISUMA , pp. 62-69.

Mendenhall, W., Reinmuth, J. & Beaver, R., 1993. *Statistics for Management and Economics*. 7th ed. Belmont, California: Duxbury Press.

- Mensah, I., Nani, G., Adjei-Kumi, T. & Adinyira, E., 2017. *Modelling Construction duration: A comprehensive review of literature*. Conference Paper, Researchgate.
- Montgomery, D. C., Runger, G. C. & Hubele, N. F., 2011. *Engineering Statistics*. 5th ed. 111 River Street, Hoboken, United States of America: John Wiley & Sons, Inc..
- Moradi, S., Nasirzadeh, F. & Golkhoo, F., 2015. A hybrid SD–DES simulation approach to model construction projects. *Construction Innovation* , 15(1), pp. 66-83.
- Mubarak, S., 2010. *Construction Project Scheduling and Control*. 2nd ed. Hoboken, New Jersey: John Wiley & Sons, Inc. .
- Myttenaere, A. d., Golden, B., Grand, B. L. & Rossi, F., 2015. *Using the Mean Absolute Percentage Error for Regression Models*. Bruges (Belgium), s.n., pp. 113-118.
- Ng, S. T. et al., 2001. , The predictive ability of Bromilow’s time–cost model. *Construction Management and Economics* , Volume 19, p. 165–173 .
- Nguyen, L. D. & Nguyen, H. T., 2013 . Relationship between building floor and construction labor productivity A case of structural work. *Engineering, Construction and Architectural Management*, 20(6), pp. 563-575.
- Nguyen, L. D., Phan, D. H. & Tang, L. C. M., 2013 . Simulating Construction Duration for Multistory Buildings with Controlling Activities. *Journal of Construction Engineering and Management*, Volume 139, pp. 951-959.
- Nkado, R. N. ., 1992. Construction time information system for the building industry. *Construction Management and Economics*, Volume 10, pp. 489-509 .
- Nkado, R. N., 1995. Construction time-influencing factors: the contractor's perspective. *Construction Management and Economics*, Volume 13, pp. 81-89.
- Odabaşı, E., 2009. *Models for Estimating Construction Duration: An Application for Selected Buildings On The Metu Campus*. Ankara, Turkey: A Thesis Submitted To The Graduate School Of Natural And Applied Sciences Of Middle East Technical University.
- Ogunsemi, D. R. & Jagboro, G. O., 2006. Time-cost model for building projects in Nigeria. *Construction Management and Economics* , Volume 24, p. 253–258.

Pardoe, I., 2012 . *Applied Regression Modeling*. 2nd ed. Hoboken, New Jersey: John Wiley & Sons, Inc .

Pawan, P. & Lortherapong, P., 2016. A Fuzzy-Based Integrated Framework for Assessing Time Contingency in Construction Projects. *Journal of Construction Engineering and Management*, 142(3), pp. 1-9.

Petruseva, S., Zujo, V. & Zileska-Pancovska, V., 2012. Neural Network Prediction Model for Construction Project Duration. *International Journal of Engineering Research and Technology* , Volume 1, pp. 1-13.

PMI, 2017. *A guide to the project management body of knowledge (PMBOK guide) / Project Management Institute*. 6th ed. 14 Campus Boulevard, Newtown Square, Pennsylvania 19073-3299 USA: Project Management Institute, Inc.

PRINCE2, 2017. *Managing Successful Projects with PRINCE2*. 6th ed. Norwich : AXELOS Limited, TSO (The Stationery Office), part of Williams Lea Tag,.

Rawlings, J. O., Pantula, S. G. & Dickey, D. A., 1998. *Applied Regression Analysis: A Research Tool*. 2nd ed. 175 Fifth Avenue, New York, NY 10010, USA: Springer.

Rezakhani, P. & Maghiar, M., 2019. Fuzzy Analytical Solution for Activity Duration Estimation under Uncertainty. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 5(4), pp. 1-13.

Sanghera, P., 2019. Project Schedule Management. In: *CAPM® in Depth*. s.l.:Cengage Learning PTR, pp. 173-220.

Sarstedt, M. & Mooi, E., 2014. Regression Analysis. In: *A Concise Guide to Market Research*. Springer-Verlag Berlin Heidelberg: Springer Texts in Business and Economics, pp. 193-233.

Sharma, G., 2017. Pros and cons of different sampling techniques. *International Journal of Applied Research* , 3(7 ) , pp. 749-752 .

Skitmore, R. M. & Ng, S. T., 2003. Forecast models for actual construction time and cost. *Building and Environment*, Volume 38, pp. 1075-1083.

Snee, R. D., 1977. Validation of Regression Models: Methods and Examples. *Technometrics*, 19 (4), pp. 415-428.

Sousa, V., Almeida, N. M. & Dias, L. A., 2014. Role of Statistics and Engineering Judgment in Developing Optimized Time-Cost Relationship Models. *Journal of Construction Engineering and Management*, pp. 1-10.

Steyerberg, E. W. et al., 2009. Assessing the Performance of Prediction Models A Framework for Traditional and Novel Measures. *Epidemiology*, 21(1), pp. 128-138.

Stoy, C., Dreier, F. & Schalcher, H.-R., 2007. Construction duration of residential building projects in Germany. *Engineering, Construction and Architectural Management*, 14(1), pp. 52-64.

Thornton, M. D., 1988 . *Construction contract durations*. Florida, USA: A Report Presented to The Graduate Committee of the Department of Civil Engineering in Partial Fulfillment of the Requirements for the Degree Master of Science University of Florida.

Tian, Y. et al., 2016 . Performance Metrics, Error Modeling, and Uncertainty Quantification. *American Meteorological Society* , Volume 144 , pp. 607-613.

Walker, D. H., 1995. An investigation into construction time performance. *Construction Management and Economics* , Volume 13, pp. 263-274 .

Waziri, B. S., 2012. *Modelling The Performance of Traditional Contract Projects in Nigeria: An Artificial Neural Network Approach*. Abuja, Nigeria, s.n., pp. 1383-1391.

Waziri, B. S., Bala, K. & Bustani, S. A., 2017. Artificial Neural Networks in Construction Engineering and Management. *International Journal of Architecture, Engineering and Construction* , Volume 6, pp. 50-60.

Yun, S. H. & Kim, S. C., 2012. Structural Work Duration Estimation and Analysis of Tower-Type Residential Construction Project. *Architectural Research*, 14(3), pp. 109-116.

Zadeh, L., 1965. Fuzzy sets. *Inform Control* , Volume 8, pp. 338-353.

Zimmermann, H.-J., 2010. Fuzzy set theory. *Advanced Review*, Volume 2, pp. 317-332.

## APPENDICES

A- 1 Structural framework construction duration survey questionnaire .....	120
A- 2 Glossary of study variables.....	121
A- 3 Case study apartment projects model building data set .....	122
A- 4 Case study apartment projects model validation data set.....	123
A- 5 Case study apartment projects construction methods used .....	124
A- 6 Statistical data of best-fit models coefficients for small-size apartments .....	125
A- 7 Statistical data of best-fit models coefficients for large-size apartments.....	125
A- 8 Predictive Performance of Simple Linear Regression (SLR) Models .....	126
A- 9 Statistical Data of the Regression Coefficients of SLR Models .....	127
A- 10 Predictive Performance of Multiple Linear Regression (MLR) Models.....	129
A- 11 Statistical Data of the Regression Coefficients of MLR Models.....	130
A- 12 Mann-Whitney U Test Statistics for Concrete mixing method .....	132
A- 13 Mann-Whitney U Test Statistics for Use of tower crane .....	132
A- 14 Mann-Whitney U Test Statistics for Shoring& scaffolding system.....	133
A- 15 Mann-Whitney U Rank and Test Statistics for Use of concrete pump .....	133

**A- 1 Structural framework construction duration survey questionnaire**

<b>Apartment project ID:</b>		<i>APat-001</i>	
Respondent position in company:			
Respondent work experience in Years:			
<b>Part 1: Apartment Project information</b>			
1.	Project name:		
2.	Project location:		
3.	Project construction start year (G.C)		
4.	Project construction end year (G.C)		
<b>Part 2: Substructure information</b>			
1.	Substructure works construction duration in days		
2.	Number of floors below ground		
3.	Isolated/combined footing <input type="checkbox"/>		
	Type of foundation: Raft/Mat foundation <input type="checkbox"/>		
	Pile foundation <input type="checkbox"/>		
<b>Part 3: Superstructure information</b>			
1.	Superstructure works construction duration in days		
2.	Number of floors above ground		
3.	Typical floor height in meter (m) units		
4.	Typical floor area in meter square (m <sup>2</sup> ) units		
5.	Type of slab: Conventional 1 way or 2 way <input type="checkbox"/> Flat slab <input type="checkbox"/> Ribbed slab <input type="checkbox"/>		
<b>Part 4: Overall Structural works information</b>			
1.	Total Building height in meter (m) units		
2.	Total number of floors		
3.	Total gross floor area in meter square (m <sup>2</sup> ) units		
4.	Total structural works construction duration in days		
<b>Part 5: Quantity of work information</b>			
1.	Quantity of excavation in meter cubic (m <sup>3</sup> ) units		
2.	Quantity of formwork in meter square (m <sup>2</sup> ) units		
3.	Quantity of reinforcing bar in kilogram (Kg) units		
4.	Quantity of concrete in meter cubic (m <sup>3</sup> ) units		
<b>Part 6: Construction methods information</b>			
1.	Concrete mixing method: On site mixing <input type="checkbox"/> Ready mix delivered with truck <input type="checkbox"/>		
2.	Use of tower crane for construction material transportation: Yes <input type="checkbox"/> No <input type="checkbox"/>		
3.	Eucalyptus wood <input type="checkbox"/>		
	Shoring & scaffolding system used for formwork: Metal <input type="checkbox"/>		
	Combination <input type="checkbox"/>		
4.	Use of concrete pump: Yes <input type="checkbox"/> No <input type="checkbox"/>		

## A- 2 Glossary of study variables

### List of dependent variables

*Substructure works duration (D\_Sub)*- the duration spent for construction of structural elements from the foundation up to and including the ground floor slab measured in days.

*Superstructure works duration(D\_Super)*- the duration elapsed for construction of structural elements starting from the ground floor slab column rising to the first floor slab up to the roof beam construction measured in days.

*Over all structural works duration (D\_Total)*- the duration took by the construction of the entire structural framework system from the foundation up to the roof beams measured in days.

### List of independent variables

No.	Identified independent variable	Unit	Symbol
1	Total number of floors	Number	<i>T_Floors</i>
2	Number of floors below ground	Number	<i>B_Floors</i>
3	Number of floors above ground	Number	<i>A_Floors</i>
4	Total Building height	m	<i>T_Height</i>
5	Typical floor height	m	<i>Ty_height</i>
6	Total gross floor area	m <sup>2</sup>	<i>T_GFA</i>
7	Typical floor area	m <sup>2</sup>	<i>Ty_area</i>
8	Quantity of excavation	m <sup>3</sup>	<i>Q_excv</i>
9	Quantity of formwork	m <sup>2</sup>	<i>Q_form</i>
10	Quantity of reinforcing bar	Kg	<i>Q_bar</i>
11	Quantity of concrete	m <sup>3</sup>	<i>Q_conc</i>
12	Type of foundation	none	<i>Found_ty</i>
13	Type of slab	none	<i>Slab_ty</i>
14	Concrete mixing method	none	<i>Conc_method</i>
15	Use of tower crane for construction material transportation	none	<i>Tower_C</i>
16	Shoring & scaffolding system used for formwork	none	<i>Form_sys</i>
17	Use of concrete pump	none	<i>Conc_pump</i>

A- 3 Case study apartment projects model building data set

ID	D_Total	D_Super	D_Sub	T_Floors	B_Floors	A_Floors	T_Height	Ty_height	T_GFA	Ty_area	Q_excv	Q_form	Q_bar	Q_conc
Apat-011	950	889	50	20	0	19	76	3.8	14,560.00	728.00	8,750.00	32,000.00	33,000.00	5,240.00
Apat-012	700	500	200	13	2	10	42	3	21,580.00	1,660.00	7,930.00	30,073.00	657,059.00	9,169.00
Apat-013	980	680	300	18	2	15	60	3	16,056.00	892.00	5,283.00	18,391.00	224,763.00	4,093.00
Apat-038	395	300	95	11	2	8	40	3	5,874.00	534.00	13,857.00	21,836.00	409,944.00	6,148.00
Apat-056	270	200	65	6	1	5	17.5	3	1,500.00	450.00	4,850.00	3,144.53	58,800.05	5,200.00
Apat-061	1,065	970	95	12	2	9	42	3	15,300.00	1,275.00	3,259.00	16,554.00	769,376.00	12,128.00
Apat-071	1,095	1,000	90	15	2	12	61.5	3.2	21,000.00	1,400.00	2,500.00	21,344.70	25,737.40	6,204.74
Apat-072	360	240	120	13	2	10	35	3.5	4,212.00	324.00	2,160.00	7,822.00	188,500.00	936.00
Apat-073	395	300	90	10	2	7	32	3	8,500.00	850.00	5,100.00	17,000.00	300,000.00	1,900.00
Apat-106	1,115	875	240	16	2	13	50	3	33,072.00	2,067.00	8,907.00	12,026.00	377,885.00	11,485.00
Apat-108	1,670	900	770	12	1	10	41	3	7,860.00	655.00	9,443.00	24,202.00	705,436.00	6,800.00
Apat-143	480	360	120	10	1	8	30	3	3,270.00	327.00	1,775.00	5,412.71	104,485.35	958.90
Apat-144	600	480	120	10	1	8	30	3	3,250.00	325.00	2,522.10	5,372.81	108,236.76	1,236.27
Apat-166	700	500	200	14	2	11	50	3.4	8,120.00	580.00	4,800.00	7,500.00	360,000.00	1,680.00
Apat-169	2,100	1,200	900	13	2	10	56.7	3.5	27,300.00	2,100.00	14,720.00	13,987.20	780,727.00	2,470.00
Apat-196	1,380	580	800	10	2	7	36	3	11,380.00	1,138.00	126.00	24,353.00	513,720.00	5,359.00
Apat-197	440	366	70	14	1	12	45	3	12,320.00	880.00	8,191.00	17,236.00	531,311.00	8,715.00
Apat-201	200	140	60	14	2	11	45	3.2	10,500.00	750.00	1,430.00	15,560.00	320,000.00	9,000.00
Apat-202	710	165	540	7	0	6	25	3	6,335.00	905.00	4,890.00	31,990.00	164,735.00	7,371.00
Apat-203	620	440	175	7	1	5	28	3	3,528.00	504.00	6,522.00	7,686.00	34,840.00	9,210.00
Apat-221	1,060	630	427	16	1	14	54	3	23,392.00	1,462.00	10,911.00	31,534.00	339,247.00	9,744.00
Apat-222	1,650	1,015	625	9	2	6	30	3	16,956.00	1,884.00	3,975.00	25,264.00	673,658.00	14,847.00
Apat-236	867	740	160	7	1	5	18.2	3	3,640.00	520.00	9,590.70	10,918.14	172,366.41	2,000.59
Apat-237	525	300	220	15	1	13	45	3	18,555.00	1,237.00	5,046.00	22,193.00	478,113.00	7,461.00
Apat-238	1,100	490	600	12	1	10	42	3	15,180.00	1,265.00	9,483.00	17,112.00	551,055.00	6,764.00
Apat-262	180	150	30	10	0	9	35	3.2	5,800.00	580.00	120.00	2,700.00	28,000.00	1,610.00
Apat-263	180	150	30	10	0	9	35	3.2	5,800.00	580.00	120.00	2,700.00	28,000.00	1,610.00
Apat-264	130	90	40	8	1	6	28.6	3.2	4,640.00	580.00	716.80	9,000.00	160,000.00	4,000.00
Apat-265	180	150	30	10	0	9	35	3.2	5,800.00	580.00	120.00	2,700.00	28,000.00	1,610.00
Apat-286	150	90	60	6	1	4	14	3	1,193.00	200.00	1,020.00	3,304.00	46,690.00	3,200.00
Apat-287	1,020	720	300	8	0	7	30	3	8,176.00	1,022.00	1,108.00	13,215.00	504,755.00	3,095.00
Apat-288	1,800	1,000	780	10	1	8	33	3	3,350.00	335.00	13,337.00	9,191.00	162,883.00	7,367.00
Apat-289	1,330	695	615	11	2	8	33	3	2,959.00	269.00	5,230.00	12,735.00	753,710.00	2,275.00

**A- 4 Case study apartment projects model validation data set**

<b>ID</b>	<b>D_Total</b>	<b>D_Super</b>	<b>D_Sub</b>	<b>T_Floors</b>	<b>B_Floors</b>	<b>A_Floors</b>	<b>T_Height</b>	<b>Ty_height</b>	<b>T_GFA</b>	<b>Ty_area</b>	<b>Q_excv</b>	<b>Q_form</b>	<b>Q_bar</b>	<b>Q_conc</b>
Apat-036	1,200	1,055	145	11	1	9	36	3	17,314.00	1,574.00	3,024.00	6,972.00	631,957.00	936.00
Apat-037	725	650	75	16	1	14	52	3	6,096.00	381.00	6,431.00	4,023.00	277,250.00	2,101.00
Apat-039	1,300	600	700	7	2	4	29	3	12,810.00	1,830.00	5,223.00	24,743.00	655,256.00	3,962.00
Apat-057	980	750	230	18	1	16	62	3	31,518.00	1,751.00	9,402.00	22,128.00	72,436.00	5,034.00
Apat-074	182	110	60	13	1	11	58.4	3.9	10,800.00	700.00	2,800.00	3,100.00	31,500.00	2,072.00
Apat-107	470	350	120	10	2	7	33	3	16,430.00	1,643.00	2,806.00	15,235.00	130,511.00	11,926.00
Apat-167	340	240	100	9	1	7	30	3	8,433.00	937.00	2,859.00	25,696.00	590,707.00	8,581.00
Apat-168	1,601	1,011	590	14	2	11	45	3	14,742.00	1,053.00	4,011.00	20,653.00	87,164.00	1,141.00

**A- 5 Case study apartment projects construction methods used**

ID	Found_ty	Slab_ty	Conc_method	Tower_C	Form_sys	Conc_pump	
Apat-011	2	2	2	2	3	1	
Apat-012	1	1	1	2	3	2	Construction method factors variables abrivations and values
Apat-013	2	3	2	2	3	1	
Apat-036	1	1	1	2	3	2	<b>Found_ty (Foundation type): (1,2,3);</b>
Apat-037	1	3	2	2	3	1	1-Isolated/combined footing
Apat-038	1	1	1	2	2	2	2-Raft/Mat foundation
Apat-039	1	3	1	2	3	2	3-Pile foundation
Apat-056	1	1	2	2	3	2	
Apat-057	1	3	2	2	3	1	<b>Slab_ty (Slab type): (1,2,3);</b>
Apat-061	1	1	1	2	3	2	1-Conventional 1way or 2way
Apat-071	2	1	1	2	3	2	2-Flat slab
Apat-072	2	2	2	2	3	1	3-Ribbed slab
Apat-073	2	2	2	2	3	1	
Apat-074	2	2	1	2	3	2	<b>Conc_method (concrete mixing method): (1,2);</b>
Apat-106	1	2	2	2	3	1	1-On site mixing
Apat-107	1	1	1	2	3	2	2-Ready mix delivered with truck
Apat-108	1	1	1	2	3	2	
Apat-143	1	2	1	2	3	2	<b>Tower_C (Use of tower crane): (1,2);</b>
Apat-144	1	2	1	2	3	2	1-Yes
Apat-166	1	3	1	2	3	2	2-No
Apat-167	1	1	1	2	2	2	
Apat-168	1	1	1	2	3	2	<b>Form_sys (Shoring&amp;scaffolding system): (1,2,3);</b>
Apat-169	2	3	2	1	3	1	1-Eucalyptus wood
Apat-196	1	1	1	2	2	2	2-Metal
Apat-197	1	3	1	1	2	2	3-Combination
Apat-201	1	1	1	1	3	2	
Apat-202	1	1	1	2	3	2	<b>Conc_pump (Use of concrete pump): (1,2);</b>
Apat-203	1	1	1	2	3	2	1-Yes
Apat-221	1	2	2	2	3	1	2-No
Apat-222	1	1	1	2	2	2	
Apat-236	1	2	1	2	2	2	
Apat-237	1	3	1	1	2	2	
Apat-238	1	1	1	2	2	2	
Apat-262	1	1	1	2	3	2	
Apat-263	1	1	1	2	3	2	
Apat-264	1	1	1	2	3	2	
Apat-265	1	1	1	2	3	2	
Apat-286	2	2	2	2	3	1	
Apat-287	1	1	1	2	3	2	
Apat-288	1	1	2	2	3	1	
Apat-289	1	1	1	2	3	2	

**A- 6 Statistical data of best-fit models coefficients for small-size apartments**

Dependent variable	Model Type		Unstandardized Coefficients		Standardised Coefficients			Collinearity Statistics	
			B	Std. Error	Beta	t	Sig.	Tolerance	VIF
<i>D_Total_s</i>	MLR	(Constant)	508.482	275.744		1.844	0.092		
		<i>T_Height</i>	-20.176	13.879	-0.358	-1.45	0.174	0.387	2.582
		<i>T_GFA</i>	0.065	0.056	0.265	1.148	0.275	0.442	2.261
	SLR	<i>Q_bar</i>	0.002	0.000	0.914	5.332	0.000	0.801	1.248
		(Constant)	224.058	89.725		2.497	0.027		
		<i>Q_bar</i>	0.002	0.000	0.831	5.387	0.000	1	1
<i>D_Super_s</i>	MLR	(Constant)	2003.84	1252.392		1.6	0.138		
		<i>Q_bar</i>	0.001	0.000	0.657	3.242	0.008	0.932	1.072
		<i>Ty_height</i>	-568.11	391.888	-0.322	-1.45	0.175	0.775	1.291
	SLR	<i>Q_conc</i>	-0.014	0.021	-0.142	-0.64	0.533	0.79	1.265
		(Constant)	190.652	64.645		2.949	0.011		
		<i>Q_bar</i>	0.001	0.000	0.706	3.596	0.003	1	1
<i>D_Sub_s</i>	MLR	(Constant)	-7.588	48.215		-0.16	0.878		
		<i>Q_bar</i>	0.001	0.000	0.659	3.868	0.003	0.602	1.661
		<i>Q_form</i>	0.014	0.005	0.546	2.956	0.013	0.512	1.955
	SLR	<i>Q_excV</i>	-0.017	0.01	-0.301	-1.66	0.125	0.531	1.883
		(Constant)	35.878	49.826		0.72	0.484		
		<i>Q_bar</i>	0.001	0.000	0.809	4.955	0.000	1	1

**A- 7 Statistical data of best-fit models coefficients for large-size apartments**

Dependent variable	Model Type		Unstandardized Coefficients		Standardised Coefficients			Collinearity Statistics	
			B	Std. Error	Beta	t	Sig.	Tolerance	VIF
<i>D_Total_L</i>	MLR	(Constant)	-187.48	356.338		-0.53	0.612		
		<i>T_Height</i>	4.943	6.323	0.101	0.782	0.454	0.99	1.01
		<i>Ty_area</i>	1.197	0.172	1.013	6.957	0	0.785	1.273
	SLR	<i>Q_form</i>	-0.029	0.013	-0.329	-2.26	0.050	0.786	1.272
		(Constant)	-285.44	223.111		-1.28	0.227		
		<i>Ty_area</i>	1.028	0.175	0.87	5.865	0.000	1	1
<i>D_Super_L</i>	MLR	(Constant)	455.168	293.351		1.552	0.155		
		<i>T_Height</i>	56.604	11.499	1.839	4.922	0.001	0.207	4.839
		<i>T_Floors</i>	-201.23	42.276	-1.738	-4.76	0.001	0.216	4.62
	SLR	<i>Q_conc</i>	0.034	0.015	0.427	2.363	0.042	0.883	1.133
		(Constant)	-40.791	185.914		-0.22	0.83		
		<i>Ty_area</i>	0.567	0.146	0.76	3.88	0.003	1	1
<i>D_Sub_L</i>	MLR	(Constant)	-139.94	153.483		-0.91	0.386		
		<i>Ty_area</i>	0.413	0.126	0.694	3.271	0.01	0.604	1.656
		<i>Q_form</i>	-0.011	0.008	-0.241	-1.3	0.227	0.783	1.277
	SLR	<i>Q_excV</i>	0.025	0.013	0.375	1.964	0.081	0.744	1.344
		(Constant)	-244.57	145.638		-1.68	0.121		
		<i>Ty_area</i>	0.459	0.114	0.77	4.009	0.002	1	1

**A- 8 Predictive Performance of Simple Linear Regression (SLR) Models**

Dependent Variable	SLR Model No.	Independent Variable	R <sup>2</sup>	R <sup>2</sup> adjusted	F-value	Sig.	Standard error of estimate	Durbin-Watson
<i>D_Total</i>	1	<i>T_GFA</i>	0.208	0.182	8.121	0.008	477.072	1.340
	2	<i>Ty_area</i>	0.260	0.237	10.916	0.002	460.893	1.275
	3	<i>Q_excV</i>	0.288	0.265	12.548	0.001	452.17	1.655
	4	<i>Q_form</i>	0.169	0.143	6.325	0.017	488.416	1.292
	5	<b><i>Q_bar</i></b>	<b>0.349</b>	<b>0.328</b>	<b>16.621</b>	<b>0.000</b>	<b>432.405</b>	<b>1.933</b>
<i>INV_D_Total</i>	6	<i>T_Floors</i>	0.123	0.095	4.366	0.045	0.00188	1.041
	7	<i>Q_form</i>	0.291	0.268	12.726	0.001	0.00169	1.099
<i>Log_D_Total</i>	8	<i>T_Height</i>	0.143	0.115	5.161	0.030	0.32372	1.070
	9	<i>T_GFA</i>	0.231	0.207	9.328	0.005	0.30653	1.151
	10	<i>Q_conc</i>	0.132	0.104	4.719	0.038	0.32571	1.108
	11	<i>Q_excV</i>	0.308	0.286	13.807	0.001	0.29081	1.639
<i>SQRT_D_Total</i>	12	<i>Ty_area</i>	0.267	0.244	11.3	0.002	8.34928	1.162
	13	<b><i>Q_bar</i></b>	<b>0.349</b>	<b>0.328</b>	<b>16.653</b>	<b>0.000</b>	<b>7.86637</b>	<b>1.775</b>
<i>D_Super</i>	14	<i>T_Floors</i>	0.132	0.104	4.717	0.038	307.304	1.371
	15	<i>T_Height</i>	0.216	0.19	8.522	0.006	292.135	1.233
	16	<i>T_GFA</i>	0.289	0.266	12.579	0.001	278.206	1.328
	17	<i>Ty_area</i>	0.285	0.261	12.331	0.001	279.001	1.278
	18	<b><i>Q_excV</i></b>	<b>0.254</b>	<b>0.23</b>	<b>10.583</b>	<b>0.003</b>	<b>284.805</b>	<b>1.557</b>
	19	<i>Q_form</i>	0.131	0.103	4.675	0.038	307.4851	1.452
	20	<i>Q_bar</i>	0.231	0.207	9.333	0.005	289.182	1.704
	21	<i>Q_conc</i>	0.123	0.095	4.342	0.046	308.927	1.265
<i>Log_D_Super</i>	22	<i>T_Floors</i>	0.169	0.142	6.302	0.017	0.30569	1.190
	23	<i>B_Floors</i>	0.129	0.101	4.596	0.040	0.31293	1.445
	24	<i>T_Height</i>	0.224	0.199	8.942	0.005	0.29542	1.032
	25	<b><i>Q_excV</i></b>	<b>0.286</b>	<b>0.263</b>	<b>12.444</b>	<b>0.001</b>	<b>0.28326</b>	<b>1.541</b>
	26	<i>Q_bar</i>	0.257	0.233	10.734	0.003	0.28901	1.490
<i>INV_D_Super</i>	27	<i>A_Floors</i>	0.130	0.102	4.638	0.039	0.00265	1.116
	28	<i>Q_form</i>	0.169	0.143	6.322	0.017	0.00259	1.345
<i>D_Sub</i>	29	<i>Ty_area</i>	0.127	0.099	4.507	0.042	255.232	1.433
	30	<i>Q_excV</i>	0.195	0.169	7.505	0.010	245.095	1.692
	31	<i>Q_form</i>	0.133	0.106	4.776	0.037	254.2715	1.238
	32	<b><i>Q_bar</i></b>	<b>0.329</b>	<b>0.308</b>	<b>15.216</b>	<b>0.000</b>	<b>223.715</b>	<b>1.971</b>
<i>Log_D_Sub</i>	33	<i>Ty_area</i>	0.513	0.497	32.633	0.000	0.00685	1.312
<i>INV_D_Sub</i>	34	<i>B_Floors</i>	0.238	0.213	9.685	0.004	0.00856	1.109
	35	<i>INV_Q_excV</i>	0.420	0.401	22.451	0.000	0.00747	1.446
	36	<i>INV_Q_form</i>	0.541	0.526	36.553	0.000	0.00665	1.186
	37	<i>INV_Q_bar</i>	0.162	0.135	6.001	0.020	0.42899	1.127

**A- 9 Statistical Data of the Regression Coefficients of SLR Models**

SLR Model No.		Unstandardized Coefficients		Standardised Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	482.143	139.035		3.468	0.002
	<i>T_GFA</i>	0.030	0.010	0.456	2.850	0.008
2	(Constant)	351.377	157.694		2.228	0.033
	<i>Ty_area</i>	0.513	0.155	0.510	3.304	0.002
3	(Constant)	440.391	128.437		3.429	0.002
	<i>Q_excV</i>	0.067	0.019	0.537	3.542	0.001
4	(Constant)	444.816	164.818		2.699	0.011
	<i>Q_form</i>	0.024	0.009	0.412	2.515	0.017
<b>5</b>	<b>(Constant)</b>	<b>405.653</b>	<b>122.548</b>		<b>3.310</b>	<b>0.002</b>
	<b><i>Q_bar</i></b>	<b>0.001</b>	<b>0.000</b>	<b>0.591</b>	<b>4.077</b>	<b>0.000</b>
6	(Constant)	0.005	0.001		3.938	0.000
	<i>T_Floors</i>	0.000	0.000	-0.351	-2.089	0.045
7	(Constant)	0.004	0.001		6.972	0.000
	<i>Q_form</i>	0.000	0.000	-0.539	-3.567	0.001
8	(Constant)	2.413	0.175		13.783	0.000
	<i>T_Height</i>	0.010	0.004	0.378	2.272	0.030
9	(Constant)	2.571	0.089		28.782	0.000
	<i>T_GFA</i>	0.000	0.000	0.481	3.054	0.005
10	(Constant)	2.603	0.103		25.247	0.000
	<i>Q_conc</i>	0.000	0.000	0.363	2.172	0.038
11	(Constant)	2.548	0.083		30.841	0.000
	<i>Q_excV</i>	0.000	0.000	0.555	3.716	0.001
12	(Constant)	18.389	2.857		6.437	0.000
	<i>Ty_area</i>	0.009	0.003	0.517	3.362	0.002
<b>13</b>	<b>(Constant)</b>	<b>19.477</b>	<b>2.229</b>		<b>8.736</b>	<b>0.000</b>
	<b><i>Q_bar</i></b>	<b>0.000</b>	<b>0.000</b>	<b>0.591</b>	<b>4.081</b>	<b>0.000</b>
14	(Constant)	131.218	188.775		0.695	0.492
	<i>T_Floors</i>	34.416	15.847	0.363	2.172	0.038
15	(Constant)	87.622	158.024		0.554	0.583
	<i>T_Height</i>	11.3	3.871	0.464	2.919	0.006
16	(Constant)	293.769	81.079		3.623	0.001
	<i>T_GFA</i>	0.022	0.006	0.537	3.547	0.001
17	(Constant)	235.812	95.46		2.47	0.019
	<i>Ty_area</i>	0.330	0.094	0.533	3.512	0.001
<b>18</b>	<b>(Constant)</b>	<b>316.439</b>	<b>80.898</b>		<b>3.912</b>	<b>0.000</b>
	<b><i>Q_excV</i></b>	<b>0.039</b>	<b>0.012</b>	<b>0.504</b>	<b>3.253</b>	<b>0.003</b>
19	(Constant)	332.203	103.762		3.202	0.003
	<i>Q_form</i>	0.013	0.006	0.362	2.162	0.038
20	(Constant)	326.807	81.957		3.988	0.000
	<i>Q_bar</i>	0.001	0.000	0.481	3.055	0.005

## A- 9 Statistical Data of the Regression Coefficients of SLR Models

SLR Model No.	Unstandardized Coefficients		Standardised Coefficients		
	B	Std. Error	Beta	t	Sig.
21 (Constant)	354.197	97.79		3.622	0.001
<i>Q_conc</i>	0.031	0.015	0.351	2.084	0.046
22 (Constant)	2.164	0.188		11.523	0.000
<i>T_Floors</i>	0.040	0.016	0.411	2.510	0.017
23 (Constant)	2.420	0.106		22.727	0.000
<i>B_Floors</i>	0.158	0.074	0.359	2.144	0.040
24 (Constant)	2.164	0.16		13.539	0.000
<i>T_Height</i>	0.012	0.004	0.473	2.990	0.005
<b>25 (Constant)</b>	<b>2.392</b>	<b>0.080</b>		<b>29.725</b>	<b>0.000</b>
<b><i>Q_excv</i></b>	<b>0.000</b>	<b>0.000</b>	<b>0.535</b>	<b>3.528</b>	<b>0.001</b>
26 (Constant)	2.404	0.082		29.353	0.000
<i>Q_bar</i>	0.000	0.000	0.507	3.276	0.003
27 (Constant)	0.006	0.001		4.356	0.000
<i>A_Floors</i>	0.000	0.000	-0.361	-2.154	0.039
28 (Constant)	0.005	0.001		5.882	0.000
<i>Q_form</i>	0.000	0.000	-0.412	-2.514	0.017
29 (Constant)	113.636	87.327		1.301	0.203
<i>Ty_area</i>	0.183	0.086	0.356	2.123	0.042
30 (Constant)	122.536	69.618		1.760	0.088
<i>Q_excv</i>	0.028	0.010	0.441	2.740	0.010
31 (Constant)	112.602	85.805		1.312	0.199
<i>Q_form</i>	0.011	0.005	0.365	2.185	0.037
<b>32 (Constant)</b>	<b>78.073</b>	<b>63.403</b>		<b>1.231</b>	<b>0.227</b>
<b><i>Q_bar</i></b>	<b>0.001</b>	<b>0.000</b>	<b>0.574</b>	<b>3.901</b>	<b>0.000</b>
33 (Constant)	1.907	0.147		12.99	0.000
<i>Ty_area</i>	0.000	0.000	0.403	2.450	0.020
34 (Constant)	0.018	0.003		6.082	0.000
<i>B_Floors</i>	-0.006	0.002	-0.488	-3.112	0.004
35 (Constant)	0.007	0.001		4.782	0.000
<i>INVQ_excv</i>	2.367	0.500	0.648	4.738	0.000
36 (Constant)	0.002	0.002		1.258	0.218
<i>INVQ_form</i>	66.377	10.979	0.736	6.046	0.000
37 (Constant)	0.004	0.002		2.816	0.008
<i>INVQ_bar</i>	556.078	97.344	0.716	5.713	0.000

**A- 10 Predictive Performance of Multiple Linear Regression (MLR) Models**

Dependent Variable	MLR Model No.	R <sup>2</sup>	R <sup>2</sup> adjusted	F-value	Sig.	Standard error of estimate	Durbin- Watson
<i>D_Total</i>	MLR1	0.472	0.436	13.391	0.000	395.9902	1.971
	MLR2	0.400	0.36	9.995	0.000	422.0296	1.782
	MLR3	0.478	0.424	8.841	0.000	400.452	1.910
	MLR4	0.473	0.418	8.67	0.000	402.3165	1.944
	<b>MLR5</b>	<b>0.504</b>	<b>0.452</b>	<b>9.815</b>	<b>0.000</b>	<b>390.3115</b>	<b>1.813</b>
	MLR6	0.488	0.435	9.197	0.000	396.662	1.875
<i>D_Super</i>	MLR7	0.359	0.316	8.401	0.001	268.454	1.711
	MLR8	0.305	0.259	6.597	0.004	279.442	1.277
	MLR9	0.397	0.357	9.896	0.001	260.272	1.439
	MLR10	0.415	0.376	10.640	0.000	256.466	1.390
	MLR11	0.371	0.306	5.694	0.003	270.544	1.651
	MLR12	0.406	0.345	6.619	0.002	262.751	1.390
	<b>MLR13</b>	<b>0.438</b>	<b>0.379</b>	<b>7.520</b>	<b>0.001</b>	<b>255.769</b>	<b>1.569</b>
	MLR14	0.436	0.378	7.479	0.001	256.073	1.525
<i>D_Sub</i>	MLR15	0.394	0.354	9.757	0.001	216.136	1.971
	MLR16	0.399	0.337	6.418	0.002	218.943	2.018
	MLR17	0.407	0.345	6.629	0.002	217.516	2.060
	<b>MLR18</b>	<b>0.428</b>	<b>0.369</b>	<b>7.237</b>	<b>0.001</b>	<b>213.569</b>	<b>2.135</b>
	MLR19	0.423	0.364	7.097	0.001	214.459	2.117

## A- 11 Statistical Data of the Regression Coefficients of MLR Models

Dependent variable	MLR Model No.	Unstandardized Coefficients		Standardised Coefficients			Collinearity Statistics		
		B	Std. Error	Beta	t	Sig.	Tolerance	VIF	
D_Total	MLR1	(Constant)	243.381	127.97		1.902	0.067		
		<i>Q_bar</i>	0.001	0.000	0.458	3.228	0.003	0.874	1.144
		<i>Q_excV</i>	0.047	0.018	0.374	2.639	0.013	0.874	1.144
	MLR2	(Constant)	266.928	147.896		1.805	0.081		
		<i>Q_bar</i>	0.001	0.000	0.445	2.641	0.013	0.705	1.419
		<i>Ty_area</i>	0.270	0.169	0.269	1.595	0.121	0.705	1.419
	MLR3	(Constant)	202.366	147.531		1.372	0.181		
		<i>Q_bar</i>	0.001	0.000	0.428	2.796	0.009	0.77	1.298
		<i>Q_excV</i>	0.044	0.019	0.351	2.347	0.026	0.807	1.239
	MLR4	(Constant)	226.596	145.980		1.552	0.131		
		<i>Q_bar</i>	0.001	0.000	0.445	2.901	0.007	0.773	1.293
		<i>Q_excV</i>	0.046	0.018	0.368	2.513	0.018	0.848	1.18
	MLR5	(Constant)	<b>143.52</b>	<b>145.659</b>		<b>0.985</b>	<b>0.333</b>		
		<i>Q_bar</i>	<b>0.001</b>	<b>0.000</b>	<b>0.35</b>	<b>2.184</b>	<b>0.037</b>	<b>0.665</b>	<b>1.505</b>
		<i>Q_excV</i>	<b>0.043</b>	<b>0.018</b>	<b>0.348</b>	<b>2.465</b>	<b>0.02</b>	<b>0.858</b>	<b>1.165</b>
MLR6	(Constant)	72.731	221.005		0.329	0.744			
	<i>Q_bar</i>	0.001	0.000	0.448	3.141	0.004	0.869	1.15	
	<i>Q_excV</i>	0.042	0.018	0.334	2.255	0.032	0.804	1.244	
MLR7	(Constant)	224.932	86.755		2.593	0.015			
	<i>Q_bar</i>	0.00044	0.000	0.346	2.212	0.035	0.874	1.144	
	<i>Q_excV</i>	0.029	0.012	0.382	2.444	0.021	0.874	1.144	
MLR8	(Constant)	178.346	158.018		1.129	0.268			
	<i>T_Height</i>	4.357	5.112	0.179	0.852	0.401	0.525	1.906	
	<i>T_GFA</i>	0.017	0.008	0.414	1.97	0.058	0.525	1.906	
MLR9	(Constant)	203.56	85.177		2.39	0.023			
	<i>T_GFA</i>	0.016	0.006	0.407	2.668	0.012	0.865	1.157	
	<i>Q_excV</i>	0.027	0.012	0.355	2.328	0.027	0.865	1.157	
MLR10	(Constant)	141.184	95.074		1.485	0.148			
	<i>Q_excV</i>	0.029	0.011	0.379	2.586	0.015	0.91	1.099	
	<i>Ty_area</i>	0.26	0.091	0.42	2.869	0.007	0.91	1.099	
MLR11	(Constant)	192.186	98.167		1.958	0.06			
	<i>Q_excV</i>	0.028	0.012	0.362	2.26	0.032	0.848	1.18	
	<i>Q_bar</i>	0.00039	0.000	0.304	1.814	0.08	0.773	1.293	
		<i>Q_conc</i>	0.011	0.015	0.12	0.734	0.469	0.808	1.238

## A- 11 Statistical Data of the Regression Coefficients of MLR Models

Dependent variable	MLR Model No.	Unstandardized Coefficients		Standardised Coefficients			Collinearity Statistics		
		B	Std. Error	Beta	t	Sig.	Tolerance	VIF	
<i>D_Super</i>	MLR12	(Constant)	121.752	150.749		0.808	0.426		
		<i>Q_excV</i>	0.026	0.012	0.344	2.221	0.034	0.854	1.17
		<i>T_Height</i>	3.195	4.835	0.131	0.661	0.514	0.518	1.929
		<i>T_GFA</i>	0.013	0.008	0.32	1.586	0.124	0.502	1.993
	MLR13	(Constant)	<b>165.956</b>	<b>87.697</b>		<b>1.892</b>	<b>0.068</b>		
		<i>Q_excV</i>	<b>0.023</b>	<b>0.012</b>	<b>0.304</b>	<b>1.977</b>	<b>0.058</b>	<b>0.819</b>	<b>1.221</b>
		<i>T_GFA</i>	<b>0.013</b>	<b>0.006</b>	<b>0.323</b>	<b>2.012</b>	<b>0.054</b>	<b>0.752</b>	<b>1.33</b>
		<i>Q_bar</i>	<b>0.00029</b>	<b>0.000</b>	<b>0.23</b>	<b>1.437</b>	<b>0.161</b>	<b>0.76</b>	<b>1.315</b>
	MLR14	(Constant)	129.694	95.563		1.357	0.185		
		<i>Q_excV</i>	0.026	0.012	0.341	2.265	0.031	0.858	1.165
		<i>Q_bar</i>	0.00023	0.000	0.179	1.045	0.305	0.665	1.505
		<i>Ty_area</i>	0.207	0.104	0.334	1.993	0.056	0.692	1.446
MLR15	(Constant)	17.917	69.847		0.257	0.799			
	<i>Q_excV</i>	0.017	0.01	0.272	1.792	0.083	0.874	1.144	
	<i>Q_bar</i>	0.001	0	0.477	3.141	0.004	0.874	1.144	
MLR16	(Constant)	35.456	79.443		0.446	0.659			
	<i>Q_excV</i>	0.018	0.01	0.286	1.827	0.078	0.848	1.18	
	<i>Q_bar</i>	0.001	0	0.504	3.08	0.004	0.773	1.293	
	<i>Q_conc</i>	-0.006	0.012	-0.078	-0.49	0.631	0.808	1.238	
<i>D_Sub</i>	MLR17	(Constant)	95.686	121.191		0.79	0.436		
		<i>Q_excV</i>	0.02	0.01	0.308	1.932	0.063	0.804	1.244
		<i>Q_bar</i>	0.001	0.000	0.487	3.172	0.004	0.869	1.15
	<i>T_Height</i>	-2.409	3.058	-0.12	-0.79	0.437	0.888	1.126	
	MLR18	(Constant)	<b>168.059</b>	<b>133.527</b>		<b>1.259</b>	<b>0.218</b>		
<i>Q_excV</i>	<b>0.02</b>	<b>0.01</b>	<b>0.322</b>	<b>2.078</b>	<b>0.047</b>	<b>0.823</b>	<b>1.215</b>		
<i>Q_bar</i>	<b>0.001</b>	<b>0</b>	<b>0.496</b>	<b>3.29</b>	<b>0.003</b>	<b>0.866</b>	<b>1.154</b>		
<i>T_Floors</i>	<b>-15.183</b>	<b>11.559</b>	<b>-0.194</b>	<b>-1.31</b>	<b>0.199</b>	<b>0.908</b>	<b>1.102</b>		
MLR19	(Constant)	137.679	120.637		1.141	0.263			
	<i>Q_excV</i>	0.02	0.01	0.319	2.049	0.05	0.822	1.217	
	<i>Q_bar</i>	0.001	0	0.474	3.145	0.004	0.874	1.144	
	<i>A_Floors</i>	-14.603	12.04	-0.177	-1.21	0.235	0.934	1.07	

**A- 12 Mann-Whitney U Test Statistics for Concrete mixing method**

Concrete Mixing method (Conc_method)		N	Mean Rank	Sum of Ranks
<i>D_Total</i>	On site mixing	29	20.4	591.5
	Ready mix delivered with truck	11	20.77	228.5
	Total	40		
<i>D_Super</i>	On site mixing	29	19.84	575.5
	Ready mix delivered with truck	11	22.23	244.5
	Total	40		
<i>D_Sub</i>	On site mixing	29	20.78	602.5
	Ready mix delivered with truck	11	19.77	217.5
	Total	40		

Test Statistics

	<i>D_Total</i>	<i>D_Super</i>	<i>D_Sub</i>
Mann-Whitney U	156.500	140.500	151.500
Wilcoxon W	591.500	575.500	217.500
Z	-.091	-.576	-.243
Asymp. Sig. (2-tailed)	.928	.565	.808

**A- 13 Mann-Whitney U Test Statistics for Use of tower crane**

Use of tower crane for construction material transportation (Tower_C)		N	Mean Rank	Sum of Ranks
<i>D_Total</i>	Tower crane used	3	12	36
	Tower crane not used	37	21.19	784
	Total	40		
<i>D_Super</i>	Tower crane used	3	11.33	34
	Tower crane not used	37	21.24	786
	Total	40		
<i>D_Sub</i>	Tower crane used	3	14.33	43
	Tower crane not used	37	21	777
	Total	40		

Test Statistics

	<i>D_Total</i>	<i>D_Super</i>	<i>D_Sub</i>
Mann-Whitney U	30.000	28.000	37.000
Wilcoxon W	36.000	34.000	43.000
Z	-1.310	-1.413	-.951
Asymp. Sig. (2-tailed)	.190	.158	.342

**A- 14 Mann-Whitney U Test Statistics for Shoring& scaffolding system**

Shoring & scaffolding system used for formwork (Form_sys)		N	Mean Rank	Sum of Ranks
<i>D_Total</i>	Metal	8	22.19	177.5
	Combination	32	20.08	642.5
	Total	40		
<i>D_Super</i>	Metal	8	20.69	165.5
	Combination	32	20.45	654.5
	Total	40		
<i>D_Sub</i>	Metal	8	24.81	198.5
	Combination	32	19.42	621.5
	Total	40		

Test Statistics

	<i>D_Total</i>	<i>D_Super</i>	<i>D_Sub</i>
Mann-Whitney U	114.500	126.500	93.500
Wilcoxon W	642.500	654.500	621.500
Z	-.457	-.051	-1.168
Asymp. Sig. (2-tailed)	.648	.960	.243

**A- 15 Mann-Whitney U Rank and Test Statistics for Use of concrete pump**

Use of concrete pump (Conc_pump)		N	Mean Rank	Sum of Ranks
<i>D_Total</i>	Pump used	10	22.05	220.5
	Pump not used	30	19.98	599.5
	Total	40		
<i>D_Super</i>	Pump used	10	23.55	235.5
	Pump not used	30	19.48	584.5
	Total	40		
<i>D_Sub</i>	Pump used	10	20.85	208.5
	Pump not used	30	20.38	611.5
	Total	40		

Test Statistics

	<i>D_Total</i>	<i>D_Super</i>	<i>D_Sub</i>
Mann-Whitney U	134.500	119.500	146.500
Wilcoxon W	599.500	584.500	611.500
Z	-.484	-.953	-.109
Asymp. Sig. (2-tailed)	.628	.340	.913