TESTING REGRESSION MODELS TO ESTIMATE COSTS OF ROAD CONSTRUCTION PROJECTS

By

SHIHUNEGN ALEMAYEHU

A Thesis Submitted to the School of Graduate Studies of Addis Ababa Institute of Technology

In partial fulfillment of the requirement for the Degree of Master of Science in Civil Engineering
(Construction Technology and Management)

November 2014

ADDIS ABABA INSTITUTE OF TECHNOLOGY

SCHOOL OF GRADUATE STUDIES

TESTING REGRESSION MODELS TO ESTIMATE COSTS OF ROAD CONSTRUCTION PROJECTS

BY: SHIHUNEGN ALEMAYEHU

APPROVED BY THE BOARD OF EXAMINERS

Name	Signature
______________________	______________________
ADVISOR	
______________________	______________________
EXAMINER (INTERNAL)	
______________________	______________________
EXAMINER (EXTERNAL)	
______________________	______________________
CHAIRPERSON	

SIGNED DECLARATION SHEET

Submitted by

Student	Signature	Date

Approved by

Advisor	Signature	Date
Chairperson,	Department' Signature	Date

Graduate Committee

Chairperson,	Faculty's Signature	Date

Graduate Committee

Dean,	Graduate School	Signature Date

CANDIDATE'S DECLARATION

I declare that this thesis entitled "**TESTING REGRESSION MODELS TO ESTIMATE COSTS OF ROAD CONSTRUCTION PROJECTS**" is my original work. This thesis has not been presented for any other university and is not concurrently submitted in candidature of any other degree, and that all sources of material used for the thesis have been duly acknowledged.

Name: ________________________

Signature: _____________________

Date: _________________________

ACKNOWLEDGEMENT

Above all, I thank the Almighty God, for what he has done to me in my entire life and make me strong and feel patience in committing the tasks in great effort to overcome various challenges and finally come up with accomplishment of this thesis.

Next, I am very grateful to express my deepest appreciation to my advisor, Assistant Professor Wubshet Jekale (Dr. Ing.), for his concern and excellent advice and also for spending his precious time to improve the quality and finally success of this research. It is with deep gratitude that I express my appreciation to Ato Mesfin Tsegaya (MSc.) and Dr:-Ing, Semu Mitku, for their advice and constructive comments in the field of study throughout. All my postgraduate instructors deserve great thanks since their inputs are worthwhile.

I would like to express my appreciation to all Ethiopian Road Authority executives and individuals who contributed directly or indirectly to this thesis and provide me with the necessary materials and support to accomplish the thesis.

SHIHUNEGN ALEMAYEHU

ADDIS ABABA INSTITUTE OF TECHNOLOGY

November, 2014

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

LIST OF TABLES

LIST OF FIGURES

ACRONYMS (ABBREVIATIONS)

AC	Asphalt Concrete
AFDD	African Development Bank
AHD	Alabama Highway Department
ANN	Artificial Neural Network
BOQ	Bid of Quantity
CIPP	Capital Improvement and Preservation Program
CPI	Consumers Price Index
CSA	Central Statistics Agency
DDS	Design Documentation Summary
DBST	Double Surface Treatment
ERA	Ethiopian Road Authority
FDRE	Federal Democratic republic of Ethiopia
FEC	Firm Estimate of Cost
Gas	Generic Algorithm
GDP	Growth Domestic Product
IPC	Interim Payment Certificate
MAPE	Mean Average Percentage Error
PAC	Preliminary Assessed Cost
PMBoK	Project Management Body of Knowledge
PS&E	Plans, Specifications and Estimate
ROC	Rough Order of Cost
ROCKS	Road Costs Knowledge System
ROW	Right Of Way
RSDP	Road Sector Development Program
SPSS	Special software for social science
US	United States
WSDOT	Washington Department of Transport

ABSTRACT

At the outset of the project, when the scope definitions are in the early stages of development, little information was available, yet there is often a need for some assessment of the potential cost. The owner needs to have a rough or approximate value for the project's cost for purposes of determining the economic desirability of proceeding with design and construction. Special quick techniques are usually employed, utilizing minimal available information at this point to prepare a conceptual estimate. Little effort is expended to prepare this type of estimate, which often utilizes only a single project parameter, such as square meter of floor area, or span length of a bridge. Using available, historical cost information and applying like parameters, a quick and simple estimate can be prepared.

The objective of this study is to develop conceptual and preliminary cost estimating models for asphalt road construction projects using historic data using statistical tools such as spss, and R-software's, based on sixteen sets of data collected in the Federal Road Projects. As the cost estimates are required at early stages of a project, considerations were given to the fact that the input data for ANNOVA F-test regression analysis to develop the cost models could be easily extracted from sketches or scope definition of the project. As a result in this study Six regression cost estimating models are developed to estimate the total cost of road construction project; among these models two include bid quantities, and four include project size (i.e. road length and road width) as input variables. The coefficient of determination (r^2) for the developed models is ranging from 0.65 to 0.98 which indicate that the predicted values from a forecast models fit with the real-life data. The values of the mean absolute percentage error (MAPE) of the developed regression models are ranging from ±16.3% for preliminary cost estimating and to ±38.9% for conceptual or ball park method of cost estimation, the results compare favorably with past researches which have shown that the estimate accuracy in the early stages of a project is between ±25% for preliminary method of cost estimating that can be related to specific characteristics of known sections or areas of the project and ±50% for conceptual method of cost estimating where early informed guesses made when virtually no drawings exist.

The research finding shows how regression models based on the significant variables or bid quantities can be used to develop regression models as tools in forecasting future road construction cost that carry much greater reliability than the previous estimated value. The paper introduces the development of cost estimating techniques and principles from historic data in the archives from both a client and consultants viewpoint both in the early stage of pre-tendering or the planning phase and project-level

Keywords: Cost estimating, Regression Model, Early cost estimate

CHAPTER ONE

1. INTRODUCTION

1.1. The Study Overview

The importance of cost models arises when it is required to estimate the unit cost of a project or total construction cost of a project in the basis of preliminary data of the project. Road cost models can be used for benchmarking to avoid cost overruns, by comparison with other similar projects. In the research described here, data have been obtained from Ethiopian Road Authority head office of each regional directorate offices at Addis Ababa. The final payment certificate contains contract amounts of bid quantities, final amount during completion of projects, and other details. The project types cover a range from gravel road construction to road upgrading, such as asphalt concrete.

The term early estimate is used to describe the process of predicting a project's cost using conceptual and preliminary methods of estimates, and usually enables the owner to determine whether to proceed with construction in the case of feasibility estimates and can reflect solutions ,identify unique construction conditions, and take into account construction alternatives using parametric techniques for determining the probable owners costs during early project development stages. The technique is used to estimate one characteristic of a system, usually its cost, from other physical and/or performance characteristics of the system. This technique involves life cycle costs, a detailed data base, and the application of multivariable correlation **[8, 41 and 42]**. In this paper a parametric estimating technique which is the least expensive, least time-consuming, least accurate than both elemental and crew development techniques were employed to address the problems of estimating during conceptual and preliminary stages of estimate.

Early cost estimating is considered as the most significant starting process to influence the a new project **[45]**. The accuracy of cost estimation improves toward the end of the project due to detailed and precise information. The early or conceptual phase is the first phase of a project process in which the need is examined, alternatives are assessed, the goals and objectives of the project are established and a sponsor is identified **[23].** At this stage, estimate accuracy is between ±25% during preliminary estimating and ±50% conceptual estimating that can helpful to designers and owners to determine whether an estimates reflects their intentions and how a change in design or construction can affect the schedule or total cost of a project due to less defined project details required to reflect a rough approximation to the cost of the project **[43]**.

Cost estimation of construction projects with high accuracy at the early phase of project development is crucial for planning and feasibility studies. Construction clients require early and accurate cost advice prior to site acquisition and commitment to build in order to enable them to take a right decision regarding the feasibility of proposed project. However, a number of difficulties arise when conducting cost estimation during the early phase. Major problems include lack of preliminary information, lack of database of works costs, lack of appropriate cost estimation methods, and the involvement of many environmental, political, social and external uncertainties. Given its significance, conventional tools such as regression analysis have been widely employed to tackle the problem.

This study presents regression models that describe the total cost of road construction project as a function of bill of quantity and project size (i.e. road length and width). The estimating models were developed based on collected data for sixteen awarded road construction projects in the Federal Road Construction projects. As these cost estimates are required at early stages of the project, considerations were given to the fact that the input data for the required model could be easily extracted from sketches or scope definition of the project.

In Ethiopia, the construction industry is one of the main economic driving sectors, supporting the national economy. It contributes to 5.3% of the GDP and public construction projects share an average annual rate of 58.2% of the Government's capital budget (Source: Sector Contribution of GDP of Ethiopia, Decb.2013). The Ethiopian construction industry continues to occupy an important position in the nation's economy even though it contributes less than the agriculture or service industries. However, the contribution of the construction industry especially road construction projects to national economic growth necessitates improved efficiency in the industry by means of cost-effectiveness and would certainly contribute to cost savings for the country through improved transport operating efficiency, access to ports and to commercial centers; access to potential resource areas; and equitable distribution among regions.

Solving estimation of project cost problem requires that decisions be made reliably and accurately towards the desired solution. Whether the problem involves a number of difficulties during cost estimation of road projects, decisions must be based on an appropriate model of the system that serves the purpose of the solution. This systematic procedure leads to decisions that are consistently reliable and reproducible. Furthermore, such a procedure can be automated for efficient processing. On the other hand, decisions made somewhat arbitrarily by engineers are open to personal interpretation and inconsistencies **[7].**

The accomplishment of the first ten years Road Sector Development Program reveals that the execution of most of the Federal road projects resulted in cost and time overruns. Studies showed that existing construction management practices exercised insufficient awareness, and time provisions to the formulation and planning phases. The construction industry suffers much either due to mis-interpretation of the actual situations or difficulty to record past project and represent these with care to project estimating practice. Becker and Behailu (2006) have also ascertained that the projects were not completed on time, within budget, and desired quality **[16]**. Cost estimation of construction projects at the early phase of project development is crucial for planning and feasibility studies.

The economic impact of the construction cost overrun is the possible loss of the economic justification for the project due to wrong advice of designers in completing with estimated cost. A cost overrun is also crucial for policy makers in their commitment for decision making in budgeting the liquid assets in a fiscal year in economic basis.

The financial impact of the cost overrun is the strain on the infrastructure projects and on national capacity in terms of foreign borrowing and domestic credits. Thus, Errors in estimating of total project costs in early project phase will have a great impact on the estimated economic rate of return of a road project.

Attempts in developing empirical estimation of cost during early project phase to minimize the effects of unrealistic cost estimation so far did little and in some cases given less emphasis in the sector. Despite all this, projects continually carried on to alleviating socio-economic problems and well being of citizens. This fact intensifies its tone when projects are public and they become parts of development objectives in the case for Ethiopia.

All discussed above made, therefore; minimizing effects of unrealistic cost estimation is one among the major task in the project management system especially during developing this in early phase of the project by consultants and clients. Above all, the efficiency and effectiveness of construction projects in cost has a great impact on the washing away of liquid assets and justified economic well being of a nation.

For the project sponsoring organization, accurate cost estimates are vital for decisions that include strategies for potential project screening, and resource commitment for further project development.

1.2. Background of the Problems

Most clients are working within tight pre-defined budgets, which are often part of a larger scheme. If the budget exceeds or quality not met the scheme could fail. Pre-contract estimating sets the original budget- forecasting the likely expenditure to the client. Therefore, the budgets resulted from estimation models should be used to positively reinforce that the design stays within the scope of the original budget.

The available cost data on road construction projects in Ethiopia was generally limited in numbers, incoherent and contain a large amount of incomplete data. Limited financial resources and lack of systematic data collection was also the major problems which affect cost data quality. These factors negatively affect to cost estimators and planners to make appropriate decisions. Accurate cost estimation in the early project development is an important issue where detailed information is not available and project cost are to be decided; in most cases cost estimation techniques are used. One of the challenges is accurate cost estimation during pre-feasibility studies.

During proposal development of this research studies, I investigate that computer estimating software in preparation of cost estimation is used by ERA to assist the controlling of elemental cost analysis developed and prepared by designer and represents the end of designer's responsibility for cost estimates. However, research studies by scholars on the cost estimation early stage of project were conducted using historic data from public road construction projects in Ethiopia is less. The primary objective of this study was examined and analyzes the cost estimation techniques with small size samples or observations in the context of Ethiopia federal road sector.

Road agencies, contractors, consultants and financial institutions need road costs information, which usually is locally available, but in many case it is scattered and collected unsystematic ways.

An approximate estimate of the cost of constructing major highway works is usually required at an early stage in the project cycle in order to determine if the scheme is reasonable and will fit in with government-funding allowance. At this early stage the proposed projects were analyzed in fundamental elements. The total road construction cost had been estimated as a function of major work activities of the project under consideration in developed cost estimation models.

The estimate was being based on an analysis of previous similar tenders using the highway client's own data or data of similar projects in Federal Road projects.

Ethiopia is one of the most promising developing countries in the sub-Saharan Africa countries and in the entire continent which shows dramatic change which scoring a sustainable level of growth. The country has experienced a massive construction program since it is apparent that infrastructure development has a direct impact on other sectors development and on a level of people's life; a well functioning construction industry plays an important role towards this end.

Increased investment on infrastructure is vital in advancing the countries productivity capacity, generating employment, in particular for the growing youth population, mitigating poverty, enhancing the country's overall economic growth and transformation, achieving greater prosperity for all citizens. Attracting investments, both domestic and foreign, should therefore be a central part of government policy and strategy. In this regard, a key objective of this study were to assist the planning of such projects to use such tools and techniques or put in place a new up-to-date techniques geared towards creating a better realistic cost estimates to plan and develop new future projects to guarantee a greater savings.

During 2001/02 to 2009/10 the Ethiopian Government spent approximately ETH birr 53 billion on construction activities **[36]**. On the other hand, whereas investment on the road project accounts nearly about 6.4 billion in the year between 1997 and 2010 in three phases. This shows that construction industry in Ethiopia has the largest economic sector in the country since 2001/02 and from this massive investment the road development program takes the lion share. Therefore, reliable estimates of construction cost by road authority and their consultants at the time of project approval are important for justifying a project on economic grounds and for planning the means of financing it. Facing with the huge economic and financial costs of expanding transport services, Government of Ethiopia is under pressure to ensure that highway projects are selected with due consideration for economic and commercial risks that arise from uncertainty in these estimates.

Ethiopian Road Authority has embarked on a coordinated road improvement work entitled "Road Sector Development Program" or (RSDP) since 1997 in three phases running through 1997-2010 with an estimated budget of US $ 1.2 and US $ 2.2 billion respectively.

The first phase of the RSDP 1 (1997 - 2002) has focused on rehabilitation and upgrading of main roads and on new construction of link and regional roads targeting to rehabilitate and upgrading about 2540km of trunk roads at an estimated cost birr 4.6 billion, (US$ 0.54 billion), upgrading/construction of 785km of link roads at birr 2.2 billion (US$ 0.26 billion), construction of 1179km new link roads at birr 0.9 billion (US$0.11 billion) and the construction of 5400km regional roads at a cost of birr 1.9 billion (US$ 0.22 billion). RSDP II envisages, the rehabilitation of 1223km, the upgrading of 2539km, the construction of 27432km, RSDP II (2002-2007) focuses on a strategy to continue the momentum to achieve the road condition targets, as well as to reinforce the process begun under RSDP I and introduces a new dimension covering the requirement of travel and transport at village level; the Ethiopian Rural Travel and Transport Sub-Program (ERTTP) **[16].**

1.3. Objective of the Study

The purpose of this study was to develop a reasonably accurate and practical method of systematized conceptual cost estimating that can be used by organizations involved in the planning and execution of road construction projects by Ethiopian Road Authority. The concepts and methodologies presented in the study can also be readily applied to other similar organizations. These should provide planners and decision makers in road agencies with tools to do early cost estimation techniques therefore, in a statistically valid and reliable manner the following objective.

- ☞ To develop early cost estimating models of road construction projects as a function of

Project size (i.e. road length and road width)

- To develop preliminary cost estimating models of road construction projects as a function of unit quantities of road construction activities

If sufficiently strong correlation can be established, then knowledge about this effect can lead to better estimates of the likely costs for future projects. In such cases there would be less measured overrun. The technique of the analysis is used multiple regression (Using the statistical micro SPSS), which allows for the simultaneous correlation of the major road works bid quantities with total road construction project costs. Such tools can simplify the use of some cost-estimating techniques and thereby facilitate rapid consideration of cost estimate alternatives.

1.4. Hypothesis

This study is based on the following hypothesis

Hypothesis:

Early cost estimation for road construction projects can be modeled using linear regression model that is:

$Y = C + b_1X_1 + b_2X_3 + \ldots + b_kX_k$

Where the symbols have the following values and meanings:

C – regression constant;

$b_1, b_2, \ldots, b_k$ – regression estimates;

$X_1, \ldots, X_k$ – independent variables;

Y – dependent variable.

Null Hypothesis:

Early cost estimation for road construction projects cannot be modeled using linear regression models.

1.5. Significance of the Study

It is obvious that all participants in road sector such as client, consultants, financer, contractors, and public should satisfy at the end results when the construction projects accomplished with optimal resources. This can be achieved by minimizing effects of unrealistic cost estimation and therefore makes this study as significant. It is necessary to ensure that the planned development of the design and procurement of a project is such that the price for its construction provides value for money (VFM) and is within the limit of anticipated by the client. Besides it facilitate the cost estimating practice relied up on realistic approach basis from project inception to completion and keeps the working environment more conducive, minimize claims and seek a best responsibility in handling the project.

1.6. Research Methodology

The research methodology is structured in theoretical and conceptual framework within the road construction by exploring of relevant topics in estimating construction costs and taking the desk study in road projects during 2007 G.C up to 2010 G.C of road sectors development program. The paper is structured as follows.

1.6.1. Literature Survey

The literature survey includes estimation practice and the theoretical and conceptual frame work in estimating costs in road projects associated to the key or critical activities under consideration. The importance of cost estimating is thoroughly investigated from literature review.

1.6.2. Data Collection

The study used secondary data sources in order to collect evidences. The study evidences were collected from project reports, final payment certificate as part of archival document instruments desk study as the research strategy. It was designed to analyze and discuss the study results of accuracy levels of each model and check for its reliability using mean average percentage error (MAPE). The research observation was taken from federal road construction projects regional office head quarter. The study projects were selected from completed road construction projects, which fulfils the information needed to track model development and mostly the completed projects investigated thoroughly for the identification of the type of projects, starting date, initial contract sum, and other primary information

1.6.3. Identification of Significant Project Cost Variables

Once the overall performance of estimating cost for road projects has been described, the analysis focuses on identifying major road construction cost as independent variables whether a significant correlation were found between them and then these variables with total project cost as dependent variable that shows a significant correlation a mathematical models can be developed. The technique of the analysis is used multiple regression (Using the statistical SPSS), which allows for the simultaneous correlation with several indicators that may themselves be correlated.

1.6.4. Result and Discussion

The analysis of the study was conducted by comparing the final project cost. The result of this apparently seen in the regression model whether the anticipated values are within the expected mean average percentage error or deviating much. A comparison made between these two values to justify the results is satisfactorily satisfying all the assumptions in estimation cost were significantly contributed for the degree of accuracy.

1.6.5. Model Development

Multiple regression analysis is used in this study to develop the parametric cost estimating model by establishing the cost estimating relationships between the road parameters and the road construction cost. Multiple regression analysis is a statistical technique used to analyze the relationship between a single dependent variable and several independent variables.

1.6.6. Evaluate Existing Approach

Regression equation and calculated values of test statistic F and the level of significance of association between the variables were computed at 5% significance level. The statistic R^2 (coefficient of determination) obtained measured how well the regression model actually fits, or its goodness of fit. If all the observation falls on the regression line, R^2 is 1. If there is no linear relationship between the dependent and independent variables, $R^2=0$. R^2 value is widely accepted to be an indicator of how well the regression model fits the population. The model usually does not fit the population as well as it fits the sample from which it is derived. The statistic adjusted R^2 attempts to more realistically reflect the goodness of fit of the regression equation in the population. The value of R^2 indicates the proportion of the variation in the dependent variable explained by the independent variable. The current treatment trend of estimating practice in estimating cost in public road project is evaluated from desk studies through Bootstrap method.

1.6.7. Regression Diagnostics

An important part of regression analysis is checking that the required assumptions are met. Residual analysis was performed to evaluate the assumptions of linearity, normality, and homoscedasticity. The residuals are the differences between the observed and the predicted values. Homoscedasticity is a description of data for which the variance of the error terms appears constant over the range of values of an independent variable (inference).

1.6.8. Conclusion

This study aims at developing early cost estimating models for road construction projects using multiple regression techniques. The models were developed based on sixteen set of data collected in the Federal Road Projects. Such types of models are very useful, especially in its simplicity and ability to be handled by calculator or a simple computer program. It has a good benefit in estimating project cost at early stages of the project since the information needed could be extracted easily from sketches or scope definition of the projects. It must be remembered that an estimated project cost is not an exact number, but it is opinion of probable cost. The accuracy and reliability of an estimate is totally dependent upon how well the project scope is defined and the time and effort expended in preparation the estimate.

1.6.9. Recommend Compatible Methods and Further Research

Recommendations to the proper relevance of estimating cost the forthcoming academicians, researchers and practitioners to enrich the requirement, implementation in future Ethiopian construction industry. Since this research was a theory oriented and initiated from the previous researches of public road construction projects it plays vital role in the improvement cost estimation practice of road construction projects. The research moreover, has forwarded theoretical intervention that can be improved the practice and result in improved outcome by applying methods used to check or review future project cost in early cost estimation practice. In accordance with the level of estimates in cost estimation process approval and review, it was also tried to show the application of statistical techniques used to forecast future project costs by the client.

1.7. Scope and Limitations of the Study

Given the objective stated above, initially my first task was to establish a sample of road construction projects what is common in this area of research, a sample large enough to allow statistical analyses of cost models. Here the first problem is that data on cost development in such projects are relatively difficult to get. The first task in model development of this study was collecting all types of road construction project data. Next, the data were clustered in such way that each group had similar properties. Such grouping was achieved by sorting all data by a predominant work activity and grouping into the similar work activity such as: asphalt overlay, reconstruction, new construction, upgrading, etc. In some cases, after clustering it was found that there were some work activities with few numbers of cases and that these cases were missed from the dataset. This study select one project type with sixteen numbers of observations. However, the available data was considered as small size sample for keeping track of multiple regression analysis. Accordingly, the study analyzed MAPE results of the developed cost estimation model of the sample and checks the problems with models through regression diagnostics using bootstraps methods.

The scope of this research is limited to the development of both a parametric cost estimating model for conceptual and preliminary cost estimating of road construction projects. The research is limited to the investigation and analysis of new road construction projects, related to state owned federal highway facilities constructed by the Ethiopian Road Authority in various regional states.

The scope of the study has been limited to the cases of Federal Road projects in RSDP. Whilst it may appear that the number of cases considered only focuses on federal Road projects, Issues related to sampling was avoided in this case by including all. This is because road projects have a similar class for this kind of analysis as they meet the criteria that it was sufficient to give an overview of comparing like with likes, provision of similar specification and similar services that would bring about better estimating practice in preliminary and early cost estimation.

1.8. Research Report Organization

This research report has been organized to comprise of the following five chapters major dissertations.

- **CHAPTER ONE- INTRODUCTION:** First it begins with describing of the *Research General Overview*. Moreover, it showed clearly *background of the problem, Objective of Research, Significance of the study* and *Research Methodology.* Finally it presented about the *Research sampling, limitation, and data collection*.
- **CHAPTER TWO- LITERATURE REVIEW:** It discussed the conceptual estimation in construction projects, project cost estimation and project development, cost estimating process, cost estimation methodology based on degree of accuracy, different early cost estimation models and crucial issues of the study.
- **CHAPTER THREE- RESEARCH METHODOLOGY:** It included *Type and Approach of the Study, Data source and Collection. First selection of forecasting model best fit to this study is determined from literature review. Next, proposed model's variables for model development are identified for the analysis. Finally, development of model based on multiple regression analysis* is done for the observation in the data collection.
- **CHAPTER FOUR- RESEARCH ANALYSIS AND DISCUSSIONS:** These parts have strongly correlated with the adjusted total cost of the project and proposed variables. After many trials a relationship that better describes the dependent variable and independent variable is taken as the best fit. Finally, the model is checked for Mean average percentage error and a regression diagnosis were made to evaluate the models.
- **CHAPTER FIVE- RESULTS AND DISCUSSIONS:** It covered the research *conclusions* on major findings of the study. The research *recommendations* covered the research investigation and contributions for improving cost estimation practice. Finally, the study forwarded some suggestions for further investigations.

CHAPTER TWO

2. LITERATURE REVIEW

2.1. Cost Estimation in Construction Projects

Estimating is the primary function of the construction industry; the accuracy of cost estimates starting from an early phase of a project through the tender estimate can affect the success or failure of a construction project. They also state that many failures of construction projects are caused by inaccurate estimates **[2]**.

In most construction project, cost estimation is used to predict the cost in advance based on a model of a project. Cost Estimate states that estimating is the primary function of the construction industry; the accuracy of cost estimates starting from early phase of a project through the tender estimate can affect the success or failure of a construction project. Since, it becomes the baseline for subsequent project cost control. If the estimate for a project is too low, the company may well lose money in the execution of the work. If the estimate is too high, the company may well lose the contract due to overpricing or cost overrun **[1]**.

A construction project has a definite budget. Therefore to achieve this, a number of methods, procedures and techniques on cost requirements have been developed to meet the set criteria in the most effective way **[24]**.

According to Hojjat Adeli and Asim Karim (2005) Cost estimates are the building block of project planning and management. Estimates are used to ascertain project feasibility and to develop and maintain detailed schedules and plans. In current industry practice, cost estimates were determined somewhat arbitrarily by estimator based on their understanding of the conditions. The objective of th is paper is briefly to describe how Cost estimation models do compared well with previous research works and show MAPE of the developed early cost estimating models.

Construction planning examines reasons for the use of cost estimation in construction projects; defie key concepts in estimation practice, effective early estimating models. The most important estimates prepared are probably for a project at sanction and a contract at tender, for it is at these points that the promoter and contractor become committed. In a project, the promoter is predicting the likely cost of completion of the works and its subsequent operation.

An approximate estimate of the cost of constructing major road construction works is usually required at an early stage in the project cycle in order to determine if the scheme is reasonable and will fit in with government-funding allowance. At this early stage the proposed project will be analyzed in fundamental elements **[22]**.

This study develop estimate models based on an analysis of previous similar tenders using the highway consultant's own data or data from client's contacts obtained from each regional team leaders final interim payment certificate and original or contract bid quantity. Therefore amount bid volume was determined from the total contract value let in each year.

Cost estimation is an essential component of infrastructure projects. Accurate estimation will assist project managers to choose adequate alternatives and to avoid misjudging of technical and economic solutions. The accuracy of cost estimation increases toward the end of the project due to detailed and precise information. The conceptual phase is the first phase of a project in which the need is examined, alternatives are assessed, the goals and objectives of the project are established and a sponsor is identified **[50]**.

Bent Flyvbjerg, Holm et al.,(2002), revealed in their study that construction cost estimating for infrastructure projects has not increased in accuracy over the past 70 years with underestimation of costs today being in the same order of magnitude that it was then. According to the study, major difficulties which arise while conducting cost estimation during the conceptual phase are lack of preliminary information, lack of database of road works costs, and lack of up to date cost estimation methods. Additional difficulties arise due to larger uncertainties as result of engineering solutions, socio-economical, and environmental issues. Parametric cost estimation or estimation based on historic database during the conceptual estimate phase is widely used in developed countries. However, developing countries face difficulties related to the creation of a road work costs database, which may be used for cost estimation in either the conceptual stage or the feasibility study of a project cycle. Even though, these difficulties are common in many countries especially in developing one, in this study much effort can be done to collect historic data which is not systematically arranged and not available in central database.

Several researchers have addressed the problem associated with cost estimation in the earlier stage of project development. Hence, cost estimates are determined utilizing experience and calculating and forecasting the future cost of resources, methods. A number of cost prediction models have been developed to salivate the anticipated cost more accurate and more reliable.

Pearce A.R.*et al.,* (1996), developed a technique for generating range estimates to evaluate the risk of cost escalation in building construction using neural networks. They outlined a procedure for building a prototype model for cost estimation, and using Artificial Neural Network (ANN) model to generate estimates for construction projects.

Hegazy T. et al., (1998) used a neural network approach to manage construction cost data and develop a parametric cost estimating model for highway projects.

Bell et al., (1987) developed multiple linear regression models for preliminary cost estimating to be used by Alabama Highway Department (AHD) for long range cost forecasting. The total project cost per mile is the function of a list of probable predictors comprising line items, such as quantities of work items per kilometer.

Mahamid et al., (2010) developed multiple linear regression models for preliminary cost estimating of road construction activities as a function of project's physical characteristics such as terrain conditions, ground conditions and soil drillability.

Early estimates, even when grossly inaccurate, often become the basis upon which all future estimates are judged, with future estimates even sometimes being ''corrected'' to be consistent with early estimates. It was found that Transportation projects have historically experienced significant cost overruns from early planning estimates.

This paper is different from those above – mentioned research works by the methods developed for small size sample through special technique called Bootstrap to evaluate the proposed cost estimation model in estimating the total cost of road construction projects. It proposes better cost estimation technique for road projects in the country and criteria for using traditional multiple regression analysis. Therefore this paper, gave alternative option for accurate project cost estimation during early feasibility or conceptual phases, and guarantees improved estimating practice for future projects by the application of statistical software through well formulated mathematical model to forecast with allowable range.

According to Schexnayder, C.J.*et al*, (2003) at an initial stage during preliminary, and conceptual, estimate accuracy is between about ±25% and ±50% respectively due to less defined project details and other uncertainties due to both internal and external factors. A typical project circle moves from the concept development phase, to the design, advertisement, bid/award, and finally the construction phase. As project progresses the accuracy of the cost estimation increases because project details becomes more clearly defined.

2.2. Cost Estimating and Project Development Level

The required levels of accuracy of construction cost estimates vary at different stages of project development, ranging from ball park figures in the early stage to fairly reliable figures for budget control prior to construction. Since design decisions made at the beginning stage of a project life cycle are more tentative than those made at a later stage, the cost estimates made at the earlier stage are expected to be less accurate. Generally, the accuracy of a cost estimate will reflect the information available at the time of estimation.

Construction cost estimates may be viewed from different perspectives because of different institutional requirements. In spite of the many types of cost estimates used at different stages of a project, cost estimates can best be classified into three major categories according to their functions. A construction cost estimate serves one of the three basic *functions*: design, bid and control **[10]**.

The time at which major cost estimation can be achieved is during planning and design for the project. Any way the cost estimation is done up to the end of the project life cycle, but the last project of the construction stage is more accurate than the beginning because of the recorded cost will be added.

Hendrickson et al. (1989) point out; a cost estimate at a given stage of project development represents a prediction provided by the cost engineer or estimator on the basis of available data. A cost estimate establishes the base line of the project cost at different stages of development of the project.

Gould (2005) defined estimate as an appraisal, an opinion, or an approximation as to the cost of a project prior to its actual construction.

According to Jelen et al., (1983) estimating is the heart of the cost engineer's work and consequently it has received appropriate attention over the years.

Holm et al., (2005), lists several reasons for making estimate, including:

- Feasibility studies
- Selection from alternate design
- Selection from alternate investment
- Appropriation of funds
- Presentation of bids and tenders

Besides the above literature, cost estimating manuals in Road transport project sectors can divide cost estimation level into four main phases or levels of project development together with the advantage explained here under:

1. Planning
2. Scoping
3. Design
4. Plans, Specifications, and Estimate (PS&E)

The estimate for each level of project development has a specific purpose, methodology, and expected level of accuracy. According to Flyvbjerg, Holm et al. 2003, explain that, while cost estimates become more accurate over time, it is exactly the cost estimate at the time of making the decision to build that is of primary interest. The major reason for the inverse relationship between the project development level and the expected accuracy range of project cost uncertainty are lack of scope definition, multiple alternatives, and lack of information about factors outside the roadway prism (right of way, community, cultural, and environmental). As the project progresses, more data is available and the expected accuracy range narrows.

For the project sponsoring organization, accurate cost estimates are vital for decisions that include strategies for potential project screening, and resource commitment for further project development. Various cost estimates are made at different stages of the project development process and include: project planning, decision to build, tendering and contracting.

This method is also the international standard for measuring the inaccuracy of cost estimates **[37]**.

2.2.1. Planning

The planning level estimate is used to estimate funding needs for long range planning and to prioritize needs for the Highway System Plan. These estimates are typically prepared with little detail to the project definition.

When using analogous project estimating, the chosen historical project must be truly analogous. Finding an appropriate project or projects and determining the similarities and differences between the historical projects and the current project can take significant time and effort. Project data from older projects is less reliable due to variations in prices, standards, construction technology, and work methods. The analogous method is best used as a tool to determine broad price ranges for simple, straight forward projects or as a check to verify estimates prepared using another method. Some of the key factors the estimator consider during his course of action.

• Due to the lack of scope definition or preliminary design, care should be taken to properly communicate with project stakeholders regarding the range of possible cost and schedule changes as the project becomes more defined.

• Given the large-scale assumptions inherent in Parametric Estimating methods, the estimator must document all assumptions clearly.

• Provide an adequate range of costs that reflects the unknowns in the project. This can be accomplished through allowances in the estimate for those items not yet defined or quantified.

• Keep the estimate current as the project waits to move on to scoping.

2.2.2. Scoping

A scoping level estimate is used to set the baseline cost for the project and to program the project. A project is programmed when it is entered into the Capital Improvement and Preservation Program (CIPP). The scoping estimate is important because it is the baseline used by the Legislature to set the budget and all future estimates will be compared against it. Clearly document assumptions and scope definitions in the Basis of Estimate document so that all future changes can be accurately compared to this estimate. Some key factors to be considered here in this level of estimate should be as follows:

• **Create/Update Basis of Estimate.** All changes, assumptions, and data origins should be clearly documented. This is particularly important because any future estimates will be compared with this one to justify changes in the cost of the project.

• Estimators should guard against false precision; that is assuming a level of precision that is not inherent to this level of estimate. Although a properly developed estimate will include well documented assumptions, many of the details that impact project cost are not defined at the time a scoping level estimate is done. Miscellaneous item allowance in design at this level of design definition typically ranges from 20 to 30 percent, and ranges even higher on nonstandard projects. This includes rounding costs (and quantities) to an appropriate significant figure.

• It is important to choose the correct unit costs for major items and then correctly inflate those costs to current dollars.

• Use sound risk identification and quantification practices to ensure that major risks to the project are identified and documented.

2.2.3. Design

Design estimates prepared at the various design levels, including Geometric Review, General Plans Review and Preliminary Contract Review are used to track changes in the estimated cost to complete the project in relation to the current budget (CIPP or "Book" amount). Each time the estimate is updated the Cost Estimate Process detailed in Figure 2.1 should be followed. The current project cost budget and schedule should be compared to the new estimate. Clearly document each of these updates in relation to the previous estimate and include the documentation in the estimate file. If the budget or scope of the project needs to be updated, fill out and submit a Project Change Request

Form. The final Engineers Estimate, along with supporting documents, is required to be filed in the Design Documentation Summary (DDS).

Design approval is an important stage of design for estimating purposes. At design approval the configuration of the project is known. This will solidify many items in the scope such as Right of Way needs, likely permit conditions, environmental mitigation, and quantities of major items and outside stakeholders. As scope definition improves, the accuracy of the estimate will likewise improve. The work effort required to prepare, document and review the estimate also increases. Some of the concerns in this level of estimate should be:

- As with the Scoping Level Estimate mentioned above, estimators should guard against false precision—thinking they know more about a project than they do. Significant project definition continues to be developed until the project is ready for advertisement. Use appropriate item allowances and ranges for estimates.
- If cost based estimating techniques are used, pay special attention to documenting all of the assumptions that are made in the development of unit prices such as the crew size, crew make up, production rates, equipment mix and type. The costs assumed for contractor overhead and profit as well as for subcontractor work should also be clearly documented. It is important to remember that these decisions may not reflect the decisions of the individual contractors that will bid the job, thus introducing elements of risk into the estimate.

2.2.4. Planning, Specification, And Estimation

The Engineer's Estimate is prepared for the Final Contract Review in preparation for advertisement and is used to obligate construction funds and to evaluate contractor's bids.

Reviews of these types of estimates should be extensive and detailed and should include final independent checks of calculations, prices and assumptions. The Basis of Estimate and overall estimate documentation package should be carefully reviewed to make sure they are complete, accurate and easily understood, and that all figures, from detailed backup to summary levels, are traceable. Some of the concerns in this level of estimate should be:

- Major quantities and cost drivers should be carefully checked to assure that they have been properly calculated (proper conversion factors have been used and allowances applied to neat line quantities if applicable).
- Specialty group estimates should be reviewed for both scope and cost.

- Contract Special Provisions should be carefully reviewed and cost and schedule impacts incorporated into the Engineer's Estimate.

2.3. Cost Estimating Process

According to **Washington State Department of Transportation** (2014), the process presented in Figure 2.1 describes the way project cost estimates developed. It is applied to all levels of project delivery, starting with the planning level estimate and ending with the project design and plan, specification, and estimate (PS&E) level. Each level of estimate may require different estimating inputs, methods, techniques and tools.

All projects benefit from following a thoughtful and deliberate process in developing project cost estimates. And also the task of cost estimating, by its very nature, requires the application of prudent judgment to the completion of the task [Washington State Department of Transportation Determine Estimate Basis].

These activities focuses on obtaining project information, including all previously developed project scope and schedule details and data, from which a project cost estimate can be prepared. The level of scope detail varies depending on the project phase, project type, and project complexity, but would include the design matrix and criteria, all assumptions and pertinent scope details. The estimate basis should be clearly documented and forms the beginning of the estimate file that should be prepared for each estimate. Each of the following steps will add information to this file, with the end result being a complete traceable history for each estimate. This documentation is covered in detail under the Documentation/Basis of Estimate section, later in this manual.

I. Prepare Base Estimate

Preparing base estimate covers the development of estimated costs for all components of a project, excluding future escalation. These components may be estimated using different techniques depending on the level of scope definition and the size and complexity of the project. The number and detail of components estimated may vary depending on the project development phase. For example, in the scoping phase the cost estimate covers preliminary engineering, Right of Way (ROW), and construction. As the design progresses and more details are known, pieces of the estimate become more detailed. Key inputs to this activity include project scope details, Historical Databases and other cost databases, knowledge of Market Conditions, and use of Inflation Rates.

A required component of the base estimate step is the preparation of a Basis of Estimate document that describes the project in words and includes underlying assumptions, cautionary notes, and exclusions. The base estimate should also be based upon, and include as an attachment for reference, the associated schedule for all remaining project activities. For conceptual level base estimates the schedule will be cursory and very broad in its coverage. However, as a minimum it should include the major milestones that WSDOT uses to measure performance and progress on projects. The conceptual level schedule may only include a few activities, but should begin with the development of the project, and include ROW, design, and construction phases.

II. Review Base Estimate

Review base estimate is necessary to ensure that (1) assumptions and basis are appropriate for the project; (2) the base cost estimate is an accurate reflection of the project's scope of work; (3) scope, schedule, and cost items are calculated properly and required components are not missing or double counted; and (4) historical data, the cost based estimate data, or other data that was used reasonably reflects project scope and site conditions. Internal specialty groups and/or Subject Matter Experts (SMEs) must participate in reviewing the Base Estimate.

III. Determine Risks and Set Contingency

Determine risks and set contingency is part of developing a risk management plan for a project, and is an integral component of project management planning – see the Project Management Online Guide. Risk management is an active and ongoing process of maximizing the probability and consequences of positive risk events (opportunities) and minimizing the probability and consequences of negative risk events (threats) to the project objectives. In the context of cost estimating, the cost impact of project risks (favorable or unfavorable) must be included to derive a total project cost.

If necessary, internal and/or external specialists are involved in a workshop format to validate the Base Estimate, provide input on specific issues such as construction staging, and elicit risks for modeling purposes. Formal or informal risk assessment techniques are a valuable and valid tool and should be applied to all estimates.

IV. Determine Estimate Communication Approach

Cost estimate data is communicated to both internal and external constituencies. The communication approach determines what estimate information should be communicated, who should receive this information, how the information should be communicated, and when the information should be communicated. Cost estimate information should be included when the communication plan is developed as part of the project management process. Often the words are as important as the numbers. The Basis of Estimate document can be used effectively as a communication tool to convey key information about the project to others.

V. Conduct Independent Review and Obtain Management Endorsement

Estimates are key products of the project management process and are fundamental documents upon which key management decisions are based. Given their importance, all estimates should receive an independent review and then be reconciled and revised as needed to respond to independent reviewer comments. Once independent review comments have been satisfactorily incorporated, estimates should be presented to management staff for approval.

Management approval of estimates developed for initial budgeting or baseline definition is a defined step in the project management process. Revised estimates, typically developed if project requirements change, or as design is developed, should also be reviewed by management staff, revised as necessary to reflect management comments, and then approved. Each revised estimates shall then be incorporated into project cost baselines through the established project change management process.

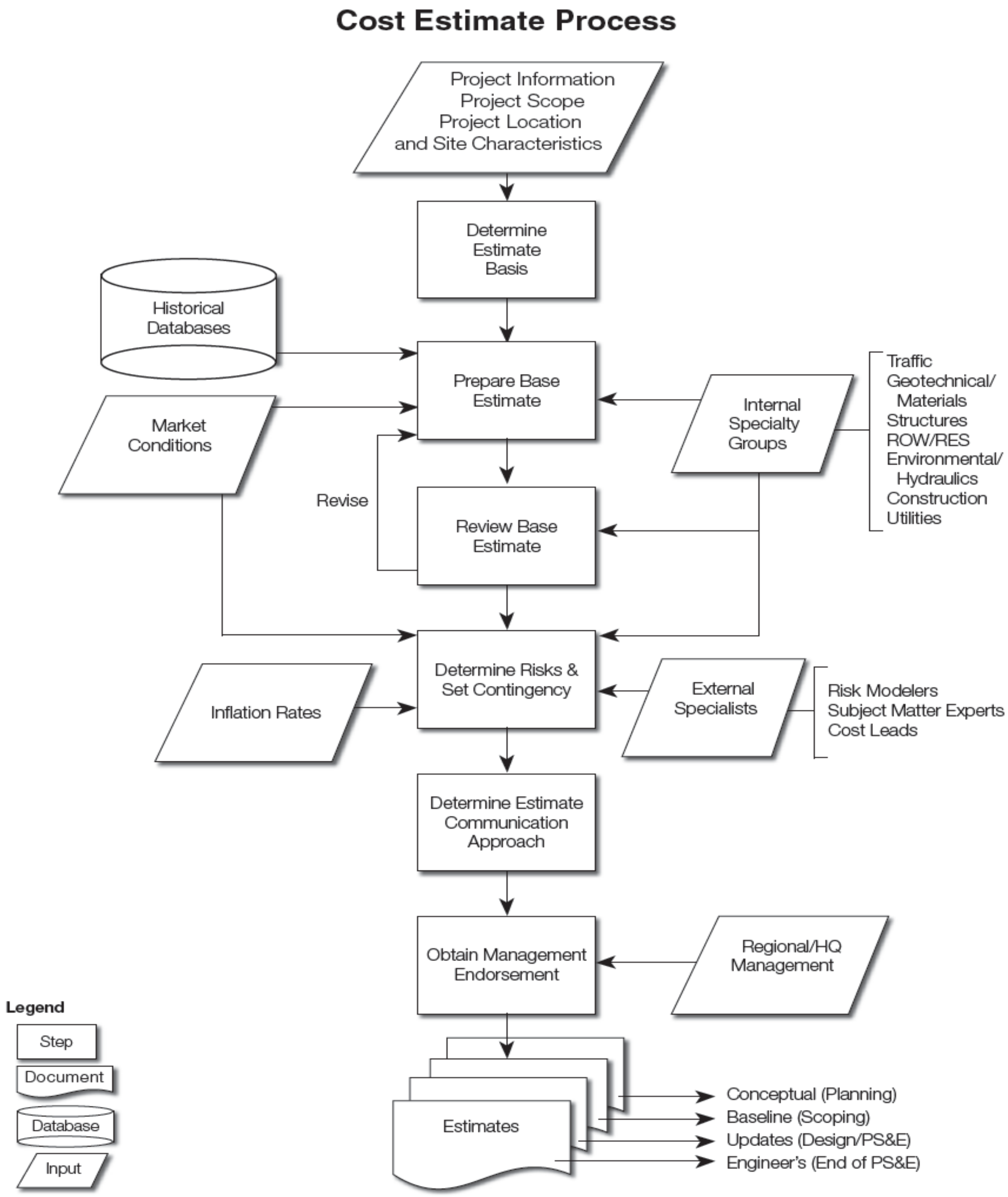


Figure 2.1 Cost Estimating Process
Source: Washington State Department of Transportation

2.4. Cost Estimating Methodology

Construction cost estimates may be viewed from different perspectives because of different institutional requirements. In spite of the many types of cost estimates used at different levels of a project, Estimating methodologies fall into one of four categories: parametric, historical bid-based, cost based and risk-based. These categories encompass scores of individual techniques/tools to aid the Estimator in preparing cost estimates. It is important to realize that any combination of the methods may be found in any given estimate.

I. Parametric methods

Parametric methods are applied to projects in the planning, scoping, or early design stage, and involve techniques that use historical data to define the cost of typical transportation facility segments, such as cost per lane mile, cost per interchange, cost per square foot, and cost per intersection. Two techniques that are commonly used in parametric estimating are (1) analogous (similar) projects and (2) historical percentages. Commercial estimating software available that can support parametric estimating for projects for which database does not have relevant historical data. Commercial estimating programs are available to assist in parametric estimating, especially for projects that have little or no historical data available in databases.

The key issue to consider when using this approach is comparing like with like: Are the standards the same in the previous projects? Does the price include infrastructure? Are professional fees and financing costs included etc.? Despite these concerns an order of magnitude estimate can be useful, particularly at the conceptual stages of projects when information is very limited and alternatives have to be ranked quickly (Norman, 1994). In this study all data were clustered in such a way that each group had similar properties. Such grouping achieved by sorting all data by a predominant work activity and grouping into similar work activity such as: asphalt overlay, reconstruction, new construction, upgrading, Rehabilitations, etc. In some cases, after clustering it was found that there were some work activities with few numbers of cases and that these cases were eliminated from the dataset. To produce the acceptable range of cost estimate during conceptual phase and propose a solution numerous trials were performed and it was a time consuming part of this study. According to Schexnayder, C. J 2003, error categorization was been set up for conceptual estimate. Proposed the error of cost estimation at the conceptual phase is approximately between ±25 for preliminary and ±50 for conceptual.

II. **Historical bid-based methods**

Historical bid-based methods are commonly used to develop Engineer's Estimates, and are appropriate when design definition has advanced to the point where quantification of units of work is possible. These methods apply historical unit costs to counts or measures of work items to determine a total cost for the item or project. The unit cost data used is typically received in bid documents from prior projects and should be modified or adjusted to reflect current prices (inflated to current time) and project specific conditions such as geographic location, quantity of item needed, and the scheduled timing of project advertisement. Techniques such as historical bid pricing, historical percentage, and cost based estimating are also used to determine unit prices.

III. **Cost-based estimate methods**

Cost-based estimate methods do not rely on historical bid data, but rather are based on determining, for an item or set of items, the contractor's cost for labor, equipment, materials and specialty subcontractor effort (if appropriate) needed to complete the work. A reasonable amount for contractor overhead and profit is then added. This method is preferable on unique projects or where geographical influences, market factors and volatility of material prices can cause the use of historical bid-based methods to be unreliable. Also, since contractors generally utilize a cost-based estimating approach to prepare bids, this method can provide more accurate and defensible costs to support the decision for contract award/rejection and to support any future price negotiations with the contractor after contract award.

Cost-based estimates require significant effort, time, and estimator experience to prepare. They should be limited to those items that comprise the largest money value of the project, typically that 20 percent of items of work that account for 80 percent of project cost. The cost of the remainder of estimate line items can be determined using Historical Bid-Based Estimate methods. This approach provides for a more efficient use of estimating resources and reduces the total time and cost of preparing Cost-Based Estimates. Cost based estimating is also a good way to check a few large items of work in a historical bid based estimate to ensure that the historical prices are still valid.

IV. Risk-based estimate methods

Risk-based estimate methods involve simple or complex analysis based on inferred and probabilistic relationships between cost, schedule, and events related to the project. It uses a variety of techniques, including historical data, cost based estimating, and the best judgment of subject matter experts for given types of work, to develop the Base Cost (the cost of the project if all goes as planned). Risk elements (opportunities or threats) are then defined and applied to the Base Cost through modeling (Monte Carlo Simulation) to provide a probable range for both project cost and schedule.

Besides, PMBOK, 2013, classifies cost estimation as Expert Judgment; Analogous Estimating; Three-Point Estimating; Bottom-up Estimating; Parametric Estimating;; Vendor Bid Analysis; Reserve Analysis; Cost of Quality in different forms of *its application* and Group Decision-Making Techniques [**Table 2.1**].

Type	Applications
Expert Judgment	✓ Guided by historical information, provides valuable insight about the environment and information from prior similar projects. ✓ Can also be used to determine whether to combine methods of estimating and how to reconcile differences between them.
Analogous Estimating	✓ Means that using the actual cost of previous similar projects as the basis for estimating the cost of the current project. ✓ Frequently used to estimate costs when there is a limited amount of detailed information about the project. Uses expert judgment. ✓ Generally less costly than other techniques, but it is also generally less accurate. ✓ It is most reliable when previous projects are similar in fact, and not just in appearance, and the persons or groups preparing the estimates have the needed expertise.
Three-Point Estimating	✓ The accuracy of single-point activity cost estimates may be improved by considering estimation uncertainty and risk and using three estimates to define an approximate range for an activity's cost: - Most likely (cM). The cost of the activity, based on realistic effort assessment for the required work and any predicted expenses - Optimistic (cO). The activity cost based on analysis of the best-case scenario for the activity. - Pessimistic (cP). The activity cost based on analysis of the worst-case scenario for the activity.
Bottom-up Estimating	✓ This technique involves estimating the cost of individual work packages or individual schedule activities with the lowest level of detail. ✓ This detailed cost is then summarized or "rolled up" to higher levels for reporting and tracking purposes. ✓ The cost and accuracy of bottom-up cost estimating is typically motivated by the size and complexity of the individual schedule activity or work package.

	✓ Generally, activities with smaller associated effort increase the accuracy of the schedule activity cost estimates.
Parametric Estimating	✓ Uses a statistical relationship between historical data and other variables to calculate a cost estimate for a schedule activity resource. ✓ Can produce higher levels of accuracy depending upon the sophistication, as well as the underlying resource quantity and cost data built into the model. ✓ A cost-related example involves multiplying the planned quantity of work to be performed by the historical cost per unit to obtain the estimated cost.
Project Management Software	✓ Such as cost estimating software applications, computerized spreadsheets, and simulation and statistical tools, are widely used to assist with cost estimating. ✓ Such tools can simplify the use of some cost estimating techniques and thereby facilitate rapid consideration of various cost estimate alternatives.
Vendor Bid Analysis	✓ In cases where projects are won under competitive processes, additional cost estimating work can be required of the project team to examine the price of individual deliverables, and derive a cost that supports the final total project cost.
Reserve Analysis	✓ Many cost estimators include reserves, also called contingency allowances, as costs in many schedule activity cost estimates. This has the inherent problem of potentially overstating the cost estimate for the schedule activity. Contingency reserves are estimated costs to be used at the discretion of the project manager to deal with anticipated, but not certain, events. These events are “known unknowns” and are part of the project scope and cost baselines.
Cost of Quality	✓ are the total costs incurred by investment in preventing nonconformance to requirements, appraising the product or service for conformance to requirements, and failing to meet requirements (rework). ✓ Failure costs are often categorized into internal and external. Failure costs are also called cost of poor quality.

Decision-Making Techniques	✓ Team-based approaches, such as brainstorming, the Delphi or nominal group techniques, are useful for engaging team members to improve estimate accuracy and commitment to the emerging estimates. ✓ By involving a structured group of people who are close to the technical execution of work in the estimation process, additional information is gained and more accurate estimates are obtained. Additionally, when people are involved in the estimation process, their commitment towards meeting the resulting estimates increases.

Table: 2.1 Types of cost estimation based on its applications [Source PMBOK, 2013].

Project cost estimates can be made at any time during the life of a Project. An estimate made near the end of a project will be more accurate than one made in the beginning, because the later estimate will contain a substantial proportion of actual recorded costs. In addition, other writers acknowledge four main estimating methods in use, varying from the very approximate to the very accurate *degree of accuracy* [**Table2.2**].

Table: 2.2 Types of cost estimation based on its degree of accuracy and applications [Source A. Lester, 2006].

Method	Degree of Accuracy	Applications
Subjective/ball park	**+/− 20% to 40%**	• Is early informed guesses made when virtually no drawings exist. • The degree of accuracy would largely depend on the experience of the estimator. • The estimator relies on his experience of similar projects. • Geographical and political factors as well as the more obvious labour and material content must be taken in to account.
Parametric	**+/− 10% to 20%**	• It would be used at the budget preparation stage, but relies on good historical data-based past jobs or experience by using well-known empirical formulae or ratios. • It can be related to specific characteristics of known sections or areas of the project. • Clearly such estimates need to be qualified to enable external factors to be separately assessed.
Comparative	**+/− 10%**	• Can be used for the preparation of the budget. • When a new project is very similar to another project recently completed, a quick comparison can be made of the salient features. • This method is based on the costs of a simplified schedule of major components which were used on previous similar jobs. • It may even be possible to use the costs of a similar sized complete project of which one has had direct (and preferably recent) experience. • Due allowance must clearly be made for the inevitable minor differences, inflation and other possible cost escalations. • The best comparative estimates are made after the work breakdown structure and task lists have been assembled, checked and coded.
Analytical	**+/− 5%**	• It may required where a contractor has to submit a fixed price tender, since once the contract is signed there be able to be no price adjustment except by inflation factors or client authorized variations. • Is the most accurate estimating method, but it requires the project to be broken down in to sections, sub sections and finally individual components. Each component must then be given a cost value (and preferably also a cost code) including both the material and labour content.

2.6. Early Cost Estimation Models

2.6.1. Introduction

Cost estimation of construction projects with accuracy at the early phase of project development is crucial for planning and feasibility studies. Construction clients require early and accurate cost advice prior to site acquisition and commitment to build in order to enable them to take a right decision regarding the feasibility of proposed project. However, a number of difficulties arise when conducting cost estimation during the early phase. Major problems include lack of preliminary information, lack of database of works costs, lack of appropriate cost estimation methods, and the involvement of many environmental, political, social and external uncertainties. Given its significance, conventional tools such as regression analysis have been widely employed to tackle the problem

Three cost prediction models were developed by Christian and Newton (1998) in order to determine an accurate cost for road maintenance. These models were developed in the province of New Brunswick based on historical data during the period 1965–1994. Based on the models and the management review, it was concluded that maintenance funding needed to be increased by 25%.

According to Han et al (2008) investigated the actual budgeting process in highway construction projects, under the research collaboration of the Korean Ministry of Construction and Transportation. They then developed two-tiered cost estimation models of highway construction projects, considering the target goals for forecasting, allowable accuracy, and available information at each phase of a project budgeting and initiation.

Previous studies show many methods which are used to forecast future construction costs. Some of these models describe construction cost as a function of factors believed to influence construction cost. The relationship between construction cost and these factors have been established from past records of construction cost. Typically, the models established in this manner have been used to estimate the cost of individual contracts. These models with their relational structure are the only models expected to provide reliable long-term estimates (Wilmot and Cheng, 2003). This study therefore, focus on developing cost estimation models that are useful in forecasting at early stage of project level. These models were developed using SPSS, regression analysis on the basis of historic data in Ethiopian Federal Road Projects during the period 2004-2008.

2.6.2. Multiple Regression Models

Regression models have been proven to be reliable and used for decades. There are some advantages of regression models such as they can be defined by mathematical expression and explain relationship between dependent variable and independent variables. Regression estimating models are widely used and have been proven to be reliable in cost estimation for decades. They are effective due to a well-defined mathematical approach, as well as being able to explain the significance of each variable and the relationships between independent variables (Sodikov, 2005). And also, regression estimation models are being able to explain the significance of each variable and relationship between independent variables. Basically, regression models are intended to find the linear combination of independent variables which best correlates with dependent variables (Lewis-Beck M. 1980).

One of the powerful techniques to overcome shortcoming of regression models is transformation dependent or independent response or both. Several types of data transformation were tested and log-log transformation was chosen for analysis. The advantage of log-log transformation lies on ease of interpretation (Carroll J et al., 1988). In this study regression analysis is also chosen to fit the developed regression models with careful consideration of the above mention problems during running the data. In this study also regression models were intended to find the linear combination of independent variables which best correlates with independent variables. The regression equation was expressed as follows:

$Y = b_0 + b_1X_1 + b_2X_3 + \ldots + b_kX_k$ ……………………… **regression model**

Where the symbols have the following values and meanings:

b_0– regression constant;

$b_1, b_2, \ldots, b_k$ – regression estimates or parameters;

$X_1, \ldots, X_k$ – independent variables;

Y – the observation of dependent variable .

The variables descriptions are mentioned in table 4.16. The goal was to estimate regression coefficients of equation. SPSS statistical tool was employed to perform regression analysis. The adjusted *R* - square value (R^2) obtained during the analysis of the data in this study were ranging from 65 to 98 for model-1 and model-6 respectively in the developed statistics. To test significance of these models ANOVA (F- test) was performed.

Levinson et al., (2003) have developed a regression model to predict the cost of new links and expansion as a function of the year of completion, duration of, and the distance from the nearest downtown. Buys et al. (2006) investigated the upgrading of the road network and expansion of overland trade in sub-Saharan Africa. They developed a cost model based on work types and country specific data.

2.6.3. Data Source and Collection

For the purpose of the investigation cost data for the first thirteen years Road Sector Development Program in three phase were obtained from Ethiopian road agencies, project bid award reports, final interim payment certificate, previously done research studies, and project completion reports, in which the data on bid estimates and actual implementation costs has been objectively checked from the Regional staff on respect to federal road projects, which were completed between 2007 and 2013. The estimating technique requires historical data. The data were collected from contracts awarded by ERA's, the clients for road construction projects, awarded in the federal road projects. The data collected comprised twenty upgrading projects from previous studies with the adjustment made to current cost, twenty-two completed projects out of these 16 are AC upgrading Projects awarded over the years 2004-2008, and ten ongoing upgrading projects. The data were tabulated to ensure that all costs were considered, none is double-counted and all are clearly defined.

2.6.4. Normalization of Project Cost

Ferry, Douglas J et.al., (1991).Cost indexes are particularly important with regards to cost analysis techniques that rely on historical or past information. A road construction project is unique undertaking. Due to its unique nature, it is extremely difficult to collect and compile large quantities of cost data for roads relating to the same reference point that is under the same time, location, and condition, in order to analyze its trends and patterns for any cost research effort. The objective of the cost index development is to measure changes in the cost of an item or group of items from one point to another. As such, the index can also be used to adjust the costs from one point to another, or to a common reference point. By bringing cost information, obtained at a number of different points, to a common base by the use of indexes, a much larger sample of data can be compiled and analyzed. This process is also referred to as the normalization of the project costs.

The construction indexes data used in this research are obtained from computation using cost indexes formulas for the subsequent years project cost for a kilometer of highway considering the cost of major cost groups such as earth work, base coarse volume, and asphalt as a cost contributors. For our assumption 2006 is taken as a base year and the unit cost index value equal to 100% and computing for coming subsequent years according to the amount of money allocated during contract award.

The indexes extracted are Means Historical Cost Indexes for national level in road construction sector.

2.6.5. Changes in Highway Costs

Table 3.1 shows the change of standard highway costs from 2006 to 2008, the rate of cost increase was substantially above the rate of inflation in subsequent years of national price index indicated in consumer price index.

2.6.6. Cost Indices

Since historical cost data are often used in making cost estimates, it is important to note the price level changes or variations over time. Trends in price changes can also serve as a basis for forecasting future costs. The input price indices of labor and/or material reflect the price level changes of such input components of construction; the output price indices, where available, reflect the price level changes of the completed facilities, thus to some degree also measuring the productivity of construction. A price index is a weighted aggregate measure of constant quantities of goods and services selected for the package. The price index at a subsequent year represents a proportionate change in the same weighted aggregate measure because of changes in prices. Let lt be the price index in year t, and lt+1 be the price index in the following year t+1. Then, the percent change in price index for year t+1 is:

Periodic changes $$i_{t+1} = \frac{I_{t+1} - I_t}{I_t}(100\%)$$

…………………………..**eq.(2.1)**

The change in index from period to period is expressed as a percentage changes as follow

From the table 3.1 for example index for the period 2008 is 140.1 and that for the period 2007 is 113.3. The index point change between these periods is equal to 26.8. Dividing this result by initial index 113.3 results index change equal to 0.2365, which is 23.65%

2.6.7. Average Periodic Change

Estimation of the future rate increase r is not at all straightforward. A simple expedient is to assume that future inflation will continue at the rate of the previous period:

Thus the average periodic change resulting from these indexes can be found by the following formula:

$$r = [(I_e/I_b)^{1/n} - 1] \times 100\%$$ ………………………**eq. (2.2)**

where,

r = average percentage change rate per period

I_e = index value at end period

I_b = index value at the beginning period

An example for r is as to the eq(3.2) using the index value in the same table above:

$r = [(I_{10}/I_6)^{1/4} - 1] \times 100\%$

$r= [(159.7/100)^{1/4} -1] \times 100\%$

$r = [1.124 -1]\times100\% = 12.4\%$

2.6.8. Forecasting Future Indexes

Future forecasts of costs will be uncertain: the actual expenses may be much lower or much higher than those forecasted. This uncertainty arises from technological changes, changes in relative prices, inaccurate forecasts of underlying socioeconomic conditions, analytical errors, and other factors. Cost indexes compares cost or price changes between periods for a fixed quantity of goods or services. Using these indexes we can obtain the cost of similar design without going through detailed estimates. A longer term perspective might use the average increase over a horizon of n past periods:

$$Cr= C_c(I_o/I_r)$$ ……………………**…eq.(2.3)**

Where,

C_c = present cost or cost of period of interest

Cr = Reference cost

I_c = index number at present or at a period of interest

I_r = index number of reference period

But to use this formula the forecast index is important during estimation of future cost of similar projects. The average rate of change of the index can be used to forecast future indexes by reformulating the equation of the rate of change. For the purpose of forecasting, it is often sufficient to project the trend of future prices by using a constant rate r for price changes in each year over a period of t years, then

More sophisticated forecasting models to predict future cost increases include corrections for items such as economic cycles and technology changes.

Ie= Ib(1+r/100)n………………..**equation (2.4)**

An example to forecast the price of total road construction cost for the period 2011. Before directly use the cost model developed in this study, first the index value for the end period 2011 should be calculated using the equation (2.4).

$I_{2011} = I_{2006} (1+12.4/100)^5$

$= 100 (1.124)^5$

=172.4%

2.6.9. Applications of Cost Indices to Estimating

In the screening estimate of a new facility, a single parameter is often used to describe a cost function. For example, the cost of a road construction work is a function of its size, and or length expressed in kilometers.

The general conditions for the application of the single parameter cost function for screening estimates are:

1. Exclude special local conditions in historical data
2. Determine new facility cost on basis of specified size or capacity.
3. Adjust for inflation index
4. Adjust for local index of construction costs
5. Adjust for different regulatory constraints
6. Adjust for local factors for the new facility

Some of these adjustments may be done using compiled indices, whereas others may require field investigation and considerable professional judgment to reflect differences between a given project and standard projects performed in the past.

An example for the application of cost indices in estimating future project at conceptual stage being the base year 2010 for the period 2011and road length 65000 meter for budgeting approval during conceptual estimate of the project cost using the formula developed in equation 5 of chapter 4 and equation (2.4) of chapter 2 of this paper.

Total project cost = $7888.25X_5$ X 172.4%

TPC = 7888.25 X 65000 X 172%

TPC = 512736250 X 172.4%

TPC = 512736250 X 172.4%

TPC = 881906350 this amount reflects the cost at completion if we compare the result with the original contract amount the difference is +212762356 which deviates by 31.7% from the real cost but it is acceptable during conceptual cost estimate of the project. Since the estimates are based on a preliminary brief, limited site information and general information about the type of construction, scope of work and possible alignment. Using the cost model developed in this study we can apply only to produce base estimate for long term financial planning that must be adjusted on the assessment of cost major risks or risk impacts to add contingency to move the base estimate towards the expected estimate. Risk assessments are used to calculate contingency and funding risks and must be identified according to clients cost estimate manuals on the basis amount deviated from past similar project costs in the same estimating stage.

Risk-based estimate methods discussed under this chapter brings with a solution in tracking cost analysis based on inferred and probabilistic relationships between cost, schedule, and events related to the project. It uses a variety of techniques, including historical data, cost based estimating, and the best judgment of subject matter experts for given types of work, to develop the Base Cost (the cost of the project if all goes as planned). Risk elements (opportunities or threats) are then defined and applied to the Base Cost through modeling (Monte Carlo Simulation) to provide a probable range for both project cost and schedule.

Another example for the application of cost indices in estimating future project at conceptual stage being the base year 2006 for the period 2011and road length 97000 meter for engineer's estimate of the project cost during preliminary stage using the formula developed in Model- 2 of chapter 4 and application of combined cost index for the coming years can be done in the following example:

Total project cost = $(89.7X_1 + 119.4X_2 + 97.6X_3)$ (index values)

Total project cost = $(89.7X_1 + 119.4X_2 + 97.6X_3)$ (1.133*1.41*1.43*1.59)

Total project cost = (89.7*789475 + 119.4*417193 + 97.6*788678) (1.133*1.41*1.43*1.59)

Total project cost = (197603724) (3.6)

Total project cost = 711375406

TPC = <u>711375406</u> this amount reflects the cost at completion if we compare the result with 770007996.72 which represents the original contract amount during contract award of the same period the difference is an amount equal to -58634590.6 which deviates by -7.6% from the real cost but it is acceptable during preliminary cost estimate of the project. An important thing we should know about in producing estimates relies on good historical data-based past projects by using a well-known empirical formulas alike to cost models developed in this study is only to forecast parametric cost whereas multiplying by the factor called future index to this calculated amount in order to obtain the budget figure on previous similar jobs.

2.6.10. Road Construction Cost Index

Road construction cost indices are measures of overall highway construction cost. They are compiled from the annual average price of a representative set of pay items in construction contracts. These indices are typically used to describe historical change in construction costs from the past to the present.

All the data were deflated to year base year 2008. Table 3.1 shows the index values over the years 2006-2008, the base year is 2008 (index = 1).The adjustment was carried out using the cost index by Federal Democratic Republic of Ethiopia central statistical agency.

The consumer price index (CPI) was taken to calculate a defilation factor in order to use past prices and quantities to track past changes in overall highway construction cost. However, if future prices and quantities could be estimated the same index used as a measure of future overall highway construction costs. The development of such an index and its use in portraying future highway costs is described in this section in table 2.3.

Table 2.3 consumer price index data value for adjusting AC projects

Year	2006	2007	2008	2009	2010	2011	2012	2013
Cost index	100	113.3	140.1	143	159.9	172.4	190	214

Source: FDRE CSA.

Since the adjustment was necessary because the projects were all tendered during the period 2004–2008 and the effect of inflation and competition had to be eliminated in order to ensure a homogenous cost comparison. In this paper since the final cost is the dependent variable the cost index during the period 2007-2012 is taken to grant price inflation. Following the adjustment, the projects were classified based on initial contract value into those ranging from $6.16x10^8$ to $1.281x10^9$ Birr and all were considered as large upgrading federal road construction projects.

2.6.11. Forecasting Method

Previous studies show many methods which are used to forecast future construction costs. Some of these models describe construction cost as a function of factors believed to influence construction cost. The relationship between construction cost and these factors have been established from past records of construction cost. Typically, the models established in this manner have been used to estimate the cost of individual contracts. These models with their relational structure are the only models expected to provide reliable long-term estimates (Wilmot and Cheng, 2003).

Regression estimating models are widely used in cost estimation. They are effective due to a well-defined mathematical approach, as well as being able to explain the significance of each variable and the relationships between independent variables (Sodikov, 2005). This study however, differ from other similar researches in that the size of the observation or sample of historic data is relatively very small and a especial method called bootstrap for handling of such case is employed to develop a more reliable estimates which is seen elsewhere.

2.6.12. Proposed Model's Variables

Many qualitative and quantitative factors affect the total road construction project cost as shown in previous studies. It is striking, however, that most of past construction models used only a few of them because the lack of information available in early stages of a project and information about qualitative factors surrounding each project are difficult to obtain.

As the objective of this study is to develop estimating models that can be handled easily using calculators or simple computer programs in the early stages of project, the models were developed based on quantitative factors that have significant impact on project cost at conceptual stage and that could be easily extracted from sketches at an early stages of a project. In this study variables significant to be used in the regression analysis were shown in following table.

Table 2.4 variable description in Federal Road Projects used in model development

Input variable	Description	Unit	Range
PWA	Predominant work activity	Category	Upgrading Asphalt Concrete
PW	Pavement Width	m	7-15
SW	Shoulder Width	m	0-3.25
EW	Earth work volume	m^3	272862-1365891
SURFCLASS	Surface Class	Category	Asphalt
SUBBASEBASEC	Sub base and Base coarse	m^3	375600-891664
AW	Asphalt Work	m^3	743115-1223549
Output Variable			
TPC	Total project cost of Asphalt concrete	Birr	$6.16x10^8$-$1.28x10^9$

2.6.13. Models Development

Once the variables to be included in the estimate equation have been identified, a series of models were developed using multiple regression analysis techniques. Regression models are intended to find the linear combination of independent variables which best correlates with dependent variables.

At project level we postulate that total project cost ($T.P.C_j$) is a function of bill of quantity such as earthwork volume, sub base and base coarse, pavement quantities, and as a function of project size (i.e. road length, road width) and the regression equation is expressed in two alternatives as follows:

Option one: As a function of bill of quantity (BOQ)

$T.P.C_j = \alpha + \sum \beta_j EWV + \sum \gamma_j BCV + \sum \lambda_j AW + \ldots + \varepsilon_j$

$T.P.C_j$ = Total cost of project type j (ETB)

EWV= Earthwork Volume (m^3)

BCV= Basecoarse Volume (m^3)

AW= pavement quantity (m^2)

RL= road length

RW= road width

Second option: Based on estimated cost of each BOQ

$T.P.C_j = \alpha + \sum \beta_j ECWV + \sum \gamma_j BCV + \sum \lambda_j AW + \ldots + \varepsilon_j$

$T.P.C_j$ = Total cost of project type j (ETB)

ECEWV=Estimated cost of Earthwork Volume

ECBCV= Estimated cost of Basecoarse Volume

ECAW= Estimated cost of pavement quantity

ECFW= Estimated cost of furniture works

A summary of the key points when estimating the cost of a project or program, the estimator needs to know more than a use of a single methods , and/or cost models or techniques and a variable for that model to develop an all inclusive (or a good) estimates. There is a substantial body of literature about dealing with good cost estimation practice, cost estimating and development level, cost estimating process, methodology, and early cost estimating models used during project cost estimation in different stages. These processes interact with each other when developing an estimate. Such knowledge can, for example, help designers, and owners to determine whether an estimate adequately reflects their intentions and to understand how a change can affect the total cost of a project. This can lead to consideration of appropriate alternative and development of better quality, realistic project cost estimation. The contribution of each of the section mentioned did little directly to this study but their consideration is essential to maintain cost effectiveness on projects. In addition, if rigorous documentation and estimating procedure is followed, the credibility of estimating will be increasing. This study recommends as estimates are key products of the project management process and are fundamental documents upon which key management decisions are based. Given their importance, all estimates should receive an independent review and then be reconciled and revised as needed to respond to independent reviewer comments.

CHAPTER THREE

3. RESEARCH DESIGN AND METHODOLOGY

Wubishet, 2004 considers research as a way of critical investigation for examining the various aspects of the problem under consideration; understanding and formulating guiding principles to govern the research procedure; and developing and/or testing theories for the enhancement of the existing situation, state or process. Hence, for the purpose of this study, research is a theory based proposing a general relationship between variables with the objective of developing mathematical models, theories and hypothesis in relation to some physical characteristic to road construction project costs or its techniques and application and/or determining resolutions to test hypothesis.

Hence, this chapter is the foundation for the achievement of the above definition of the study applicability. With the above research definition of the study as a foundation, it begins with the development of the study type and approach. On the bases of the study type and approaches, the chapter were continues its explanation on the study samples and populations and its selection techniques for the study used instruments for evidences collection. Finally, it covered the considerations in the study data collection and management for results analysis and discussion.

3.1. Type and Approach of the Study

This research investigation was descriptive and inferential in nature and contributes to the body of knowledge by expanding previous research investigations on analogous and parametric cost estimating for road construction projects in Ethiopia. It is descriptive and inferential because, it tried to describe the actual relationships of historic cost data for cost model development and prove the relationship between parameter and statistics. The identification of road parameter, as a significant cost driver for road construction cost. The identification of this significant road parameter indicates the importance of this parameter. The introduction of bootstrap method for assumptions of linearity, normality, and homoscedasticity for small size samples in the parametric cost estimating model developed in this study is shown to significantly improve the cost estimating model performance. For the purpose of this thesis, research is defined as the systematic and scientific investigation of quantitative properties and phenomena and their relationships. The objective of this type of research is to develop and employ mathematical models, theories and hypotheses pertaining to real life problems. In this study the process of measurement is central to quantitative research because it provides a fundamental connection between empirical observation and mathematical expression of an attribute. It starts with a theory or general statement proposing a general relationship between variables. In this approach the purpose is to take an objective position and to treat phenomena as hard and real. Research can either be a theory based (deductive) initiated for theory contribution (inductive) approach to research. This research identifies road cost parameters, as a significant cost driver for road construction cost and developed a realistic estimation model.

The type and approach of this study were categorized in consideration of research classification systems for research methodology and design application (Figure: 3.1). According to Wubishet, 2004 edited on RESEARCH METHOD for construction by Richard Fellows and Anita Liu, 2008 research can be classified:

- Based on purpose of research: *Basic* (for theoretical problems) and *Applied* (for practical problems);
- Based on the specific objectives of the research: *Descriptive* (developing facts about problems), *Exploratory* (to test, or explore, aspects of theory) *and Explanatory* (to answer a particular question or explain a specific issue) and *Co-relational* (makes comparisons, looking for trends or tendencies); and
- Based on approaches of research: *Qualitative* (to gain insights on concerning issues) and *Quantitative* (concerned with the measurement of issues).

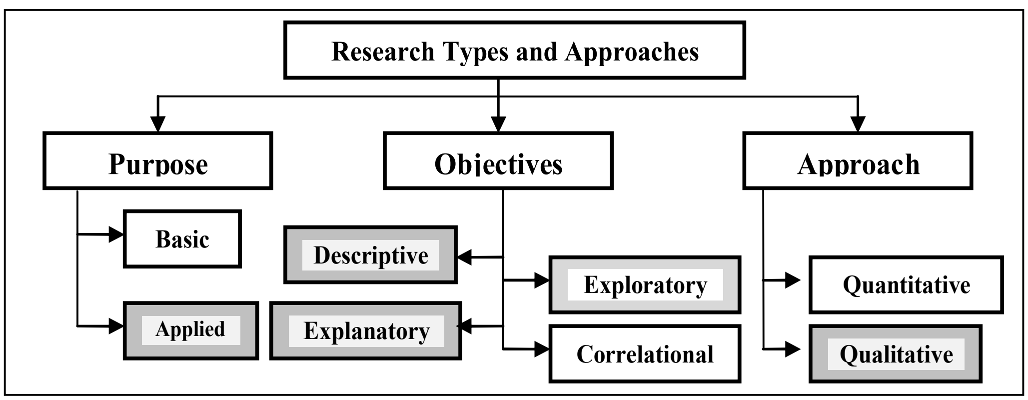


Figure 3.2: Types and approaches of a research[2] for research methodology and design [Source: Wubishet, 2004]

Since the study motivated through research conducted by scholars to prove the estimating accuracy level of early cost estimation of Federal Road Construction Projects is within the range of acceptable level to confirm with previous researches. Therefore, the study develops research hypothesis; research objectives and research problem and [refer chapter one introduction].

3.2. The Methodology Approach

Empirical estimation of cost functions requires statistical techniques which relate the cost of constructing or operating a facility to a few important characteristics or attributes of the system. The role of statistical inference is to estimate the best parameter values or constants in an assumed cost function. Usually, this is accomplished by means of regression analysis techniques.

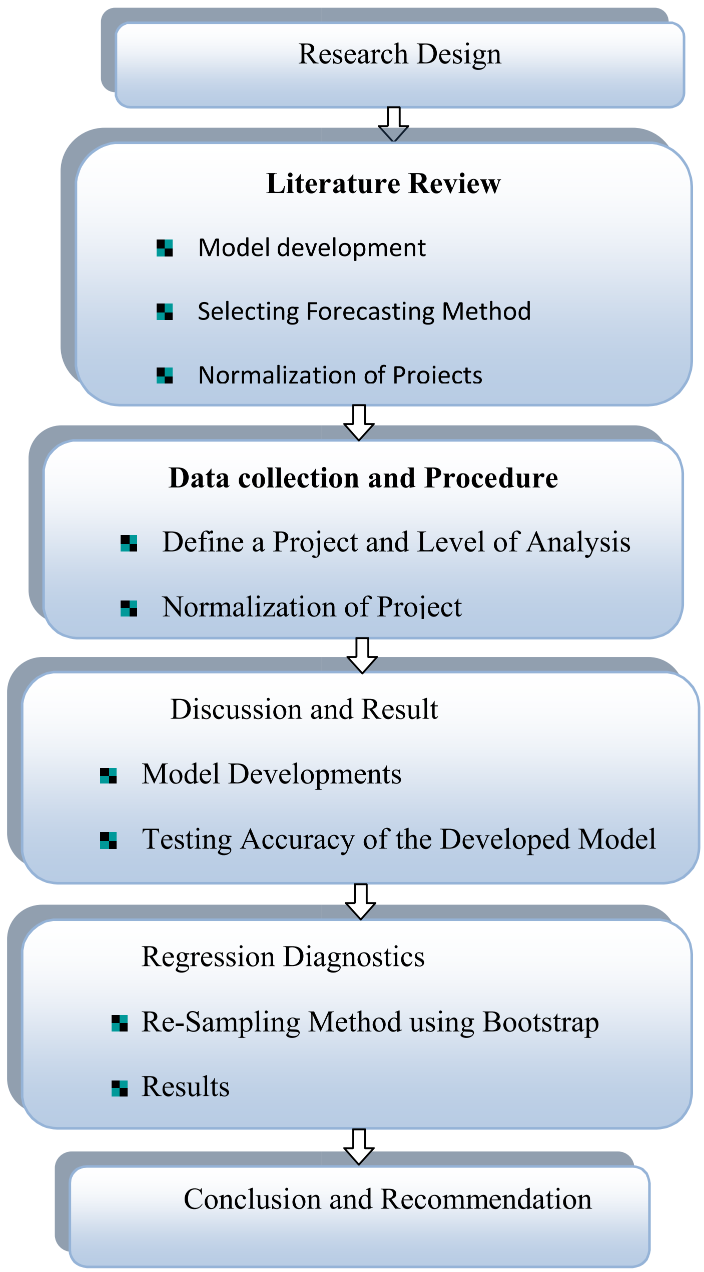


Figure 3.1 **Flowchart showing the methodology of research**.

3.3. Study Population and Sampling

In this study, the concept of research population was understood as the set of Road construction projects from which the research samples are drawn. Therefore, selection of an appropriate research population controls extraneous variation and helps to define the limits for generalizing the study findings. This study uses two (2) major asphalt road construction projects such as on the completed public road construction projects in Federal Road Projects as the study population. Those are *rehabilitation and construction.* Hence, those are parents of this study throughout the finding processes and report writing.

Projects types were selected based on their class categories fulfilling the criteria that the required information for running analysis using software is completed. However, class and type of categories were not the basic requirements of the study findings. It was mainly acquired in consideration of information they contribute in relatively big projects, comparatively better access of their physical characteristics and other related inform in higher degree than the others.

As a means of obtaining the population samples the study used non-random sampling *(Snowball)* techniques. Snowball samplings were involved in data which are difficult to access and when the inclusion of all in the analysis results better accuracy quantitative research as of this study.

3.4. Research Design and Instruments

Research design is the critical consideration for determining the most appropriate methodology and method/s. It links the data collection and analysis to yield results, and, hence, conclusions, to the research objective. The main priority is to ensure that the research maximizes the chance of realizing its objectives. Research designs were needed to take a more comprehensive view of context in order to locate units of study more precisely in relation to the significant variables that helps in developing cost models.

Hence, in due consideration of the empirical information availability, advantages of the method used and approach of the study (for both research problems and objective) the study used secondary data sources in order to collect evidences. The study evidences were collected from project reports, final payment certificate as part of archival document instruments desk study as the research strategy.

It was designed to analyze and discuss the study results of previous research works on accuracy level of each model and check for its reliability using mean absolute percentage error (MAPE). Accordingly, archival documents analysis was the research designs discussed hereunder for study major finding investigations.

3.4.1. Archival Documents Analysis

The archival documents were taken from completed public building construction projects and investigated thoroughly in order for obtaining the information related to major road construction activities and associated cost. In this study archival documents was includes evaluation reports, final payment certificate, statistical tools such as ANOVA F-test, r- software interpretation related documents, similar studies and other records related to the study concepts.

3.5. Data Collection and Management

Data collection is generally difficult, and when it comes to Developing Countries, it is worst and tiresome [Wubishet, 2004]. This was also the case in this study as well. During my data collection two major hindrances were revealed like Wubishet, 2004 listed. The first challenge was poor documentation, and Fragmentation of accessible data (Low awareness of the uses and benefits of documentation by the practitioners, Untrained and unskilled manpower usage in documentation, and Lack of sufficient modern information technology and management resources had been observed to contribute to these causes) and access to data were generally very difficult and biased; and hiding information was practiced as an instrument for resource or expert power. Further to these, misuse in the very benefit of evaluations by Governments of Developing Countries resulted in practitioners' fear of detrimental measures.

The primary aim of data collection for the study was to maximize the accuracy of developed cost model of the study in using instruments. Since this study used secondary data sources archival documents analysis.

In accordance with the study evidences collection survey activities, the projects were clustered in many types as part of the analysis for results and discussions. Consequently, asphalt road construction projects chosen and observation with survey were hierarchy in Microsoft Excel spreadsheet software. Finally regression analysis using SPSS ANOVA t-test in mathematical cost model development to achieve the objective results and

To test the accuracy level of the developed model regression diagnostic test performed using Re-Sampling method using Bootstrap techniques to arrive conclusion and recommendation of this study. All the study necessary evidences were collected from October 1st to April 1st. The essence of all data sources was directly related to the objective of the study-***developing early cost estimating models of road construction projects as a function of project size: to develop preliminary cost estimating models of road construction projects as a function of quantities of major activities***. After collecting all the available documents immediately the research analysis and discussion of results were conducted (chapter four of the research report dissertations).

Therefore, analysis and discussions of the study results dissertation in the next chapter constituted the two basic concerns of findings in project early cost model development. Finally, the study major finding was covered as improving cost estimation practice in asphalt road construction projects. It constituted two parts: developing the study reality and testing accuracy of developed model. Also, it was designed in accordance with the study results, review of related findings and developmental approaches in the study knowledge area.

HAPTER FOUR

4. RESULTS AND DISCUSSION

4.1. Correlation of Variables in Road Projects

The correlation between dependent variables and actual and estimated costs is the focal point the research to precede further explanation in the regression analysis. The squared correlation between the actual and estimated outcomes summarizes the overall accuracy of the estimation. If the estimates were no better than guesses made at random, the correlation should not be significantly greater than zero; while if the estimates were such that they either predicted all out-comes exactly (or were the multiple of the actual outcome), then the correlation would be one hundred percent.

Table 4.1: Correlation between variables in projects.

Project type	Variables	Correlations
ALL	Contract cost-Final cost at completion	0.78
AC	Contract cost-Final cost at completion	0.74
AC	Earth work and basecoarse	0.049
AC	Earth work volume and asphalt	0.17
AC	Base coarse and asphalt	0.22

In fact the squared correlation between the actual and estimated costs for all upgrading projects in the 40 collected historic data in all project type and asphalt project were 78%, and 74% respectively. Cost values significantly greater than zero, as expected, allowing as rejecting the null hypothesis that estimated values are no better than random guesses. The lower the correlation the more likely it appears that any estimate chosen at random (for a new project) will diverge by the given amount from the actual outcome.

During screening of projects only one project with truly exceptional case in the hierarchy in the project group's of the 16 asphalt upgrading projects cost model development. Estimate values for bid quantities and total project cost reveals inconsistency. Regression analysis excludes projects observation during testing individual variables for the outliner. After estimating a model, checking the entire regression for normality of residuals, omitted and unnecessary variables were reviewed.

4.2. Early Cost Estimation Model Development

The parametric method in conceptual cost estimating involves the identification of the significant road parameters, or cost drivers, and the development of the parametric model or equation. The parametric model is a mathematical model or function, which defines the cost estimating relationships (CERs). Multiple regression analysis is used in this study to develop the parametric cost estimating model by establishing the cost estimating relationships between the road parameters and the road construction cost. Multiple regression analysis is a statistical technique used to analyze the relationship between a single dependent variable and several independent variables.

Once the correlation between estimates and actual cost for road upgrading projects has been described, the analysis focused on cost model development. After a set of probable predictors were identified for each project, a statistical model was developed using multiple regression technique. Though, the small sample size was the main issue to accurately describe the parameters of the total project cost in this research paper, however using bootstrap statistical method we are able to handle the problem. The detail about the bootstrap re-sampling procedure about regression is discussed in the section 4.5.1. Finally, the equation is best fit to be used the most significant variable calibrated and taken as a final. Given the quantities per project for the predictor variables, the regression model predicted the total cost in Ethiopian Birr for that project, based on statistically significant variables as shown in Table 4.1, where the p-value for all coefficients considered in the model are less than or equal to 0.05: in this study the two cost estimation model developments based on the following major variables:

I. Estimate Based On Engineer's Bill Of Quantity
II. Estimate Based On Project Size

4.2.1. Estimate Based on Engineer's Bill of Quantities

The engineer's estimate is based on a list of items and the associated quantities from which the total construction cost is derived. This same list is also made available to the bidders if unit prices of the items on the list are also solicited from the bidders. Thus, the itemized costs submitted by the winning contractor may be used as the input variable for model development for total road construction project cost. The corresponding unit prices of the work items on the list are also used to develop models in a similar way that of bill of quantities. Hence, the estimate based on the engineers' list of quantities for various work items essentially used to estimate the total cost of road construction project in Ethiopian birr as follows:

a) Total project cost as a function of major road construction activities such as earth work, base coarse, and pavement
b) Total project cost as a function of equivalent estimated costs of major road construction activities during bid award including furniture cost.
c) Total project cost as a function asphalt quantity

Accordingly the following models were developed as shown in table below.

Table 4.2: Cost Estimation Models Developed

1. Estimate based on BOQ			
Type	Model #	Description	Remark
1.1.	1	Earth work volume(m^3), sub base and base coarse volume(m^3) and asphalt or pavement(m^2)	Significant without intercept
1.2	2	Earth work cost, sub base and base coarse cost and asphalt or pavement cost	Alternative to model-1 but it has lesser value of MAPE and (r^2) model-1 has to be more preferable to use.
2. **Estimate based on project size (i.e. length and width**			
2.1.	3	asphalt or pavement(m^2)	Significant without intercept
2.2.	4	Road width and Length	Since road with is variable along the entire length only road length is significant with *TPC .
2.3	5	Road size	Significant without intercept but it is the product of road width and length
2.4	6	Individual significant variables with road size and then TPC as a summation of this individual costs.	Individual variables were Significant without intercept

Note: * TPC is total project cost

4.2.2. Cost Estimates Based on Project Size

Screening cost estimates are often based on a single variable representing the capacity or some physical measure of the design such as floor area in buildings, length of highways, and volume of storage bins and production volumes of processing plants.

The general conditions for the application of the single parameter cost function for screening estimates are:

- Exclude special local conditions in historical data
- Determine new facility cost on basis of specified size or capacity
- Adjust for inflation index
- Adjust for local index of construction costs
- Adjust for different regulatory constraints
- Adjust for local factors for the new facility

Some of these adjustments may be done using compiled indices, whereas others may require field investigation and considerable professional judgment to reflect differences between a given project and standard projects performed in the past.

In estimating the total road construction cost of a highway project with a similar specification for a know length and width of the road with in similar economy of scale.

In this research, the following two cost estimation models base on project size namely length and road width were used.

Table 4.3: Multiple regression among total project cost and activities' quantity

Regression statistics			**Variables**	**Coef.**	**P-value**
			intercept	8.505E7	0.35
R Square	0.87		Earth works(m^3)	81.52	0.014
Adjusted R Square	0.84		Basecoarse (m^3)	620.60	0.0005
Observations	16		Pavement (m^2)	409.3	0.003
F	25				

At first, the resulting linear regression model appears to be reasonable. The model summary reports an R^2 of 0.87. The P- values for the estimates of the Earthwork volume, Base coarse volume, and Asphalt work/Pavement quantity coefficients are less than 0.05, indicating that their effects are not due to chance. Somewhat, the p-value of the constant is 0.35 which is higher than 0.05, meaning it is not significant to be included in the model. As a result, the model uses other variables as the input variable is formulated.

Table 4.4: multiple regression among total project cost and quantity without intercept

Regression statistics			**Variables**	**Coef.**	**P-value**
			Earth works(m^3)	**89.70**	**0.006**
R Square	**0.98**		**Basecoarse (m^3)**	**119.4**	**0.000**
Adjusted R Square	**0.98**		**asphalt work(m^2)**	**97.6**	**0.001**
Observations	**16**				
F	**208.5**				

Total cost (ETB) = $89.7X_1 + 119.4X_2 + 97.62X_3$..................................(Model 1)

Where,

X_1 = Earth work ; cut, fill, and topping quantities (m^3)

X_2 = Sub base and Base coarse quantity (m^3)

X_3 = Pavement quantity (m^2)

Data analysis show that the earth works, base works and asphalt works are the major activities in the 16 examined road projects. Figure 4.1 shows the average contribution of these activities in the total project cost.

Figure 4.1: Average contributions of major road construction activities in the total project cost

As the furniture cost includes a lot of work items, an equation that describes the total project cost as a function of quantities of major construction activities excluding furniture work is formulated.; this makes the model more easy to be use in the early stages where a little information are available. The coefficient of determination (r^2) of the developed model 0.98. The analysis of variance test confirmed the statistical significance level of 0.05 as shown in the table 3.5.

Where,

X_1 = Earth work (cut, fill, and topping) quantities (m^3)

X_2 = Sub base and Base coarse quantity (m^3)

X_3 = Pavement quantity (m^2)

Three individual regression models that describe the cost of earth works, basecoarse works and asphalt works as a function of bid quantities are developed. The results are shown [in **Table 4.5**]. The analysis of variable test confirmed the statistical significance of the models at a significance level of 0.05.

Table 4.5: Regression models among cost of each major activity in road construction and its quantity.

Activity	**Variables**	**Coef.**	**r^2**	**P-value**	**F**	**Best fit model**
Earthworks	Earthwork quantity (m^3) =X_1	45.79	0.76	≈ 0	44.8	Earthworks cost =45.79 X_1
Base works	Base works quantity (m^3) = X_2	151.41	0.83	≈ 0	70.10	Base works = 151.4X_2
Asphalt works	asphalt quantity (m^2) =X_3	159.17	0.86	≈ 0	87.60	Asphalt works = 159.17X_3

The furniture cost can be estimated using one of the following equations (**Table 4.6**):

Table 4.6: Cost estimation model for furniture works

Equation #	**Variables**	**P-value**	**F**	**r^2**	**MAPE**	**Best fit model**
1	Road size(m^2) = X_4* X_5	0.000	25.82	0.65	83.53	Road Furniture cost = 23.37 X_4* X_5
2	Pavement (m^2) =X_3	0.001	52.7	0.79	62.57	Road Furniture cost = 36.07X_3

Table 4.6 shows that Equation2 is the best to be used for furniture cost estimation as it has the highest r^2 and the lowest MAPE. The coefficient of determination r^2 for model 2, using equation 2

for road furniture cost estimation (shown in table 4.6), and MAPE were 45.Then the developed model is:

Total project cost =45.79 X_{1} + 151.4X_{2} +159.17X_{3} +36.07X_{3}

Total project cost =45.79X_{1} + 151.4X_{2} +195.24X_{3} ………………………………………… (Model 2)

Model 2 is similar to model 1, but model 2 could be used to estimate cost of individual construction activity (i.e. asphalt cost, base coarse cost, earthworks cost, and furniture cost).

It can be seen that model 1 is better for estimates of the total project cost because it has better accuracy (MAPE) than model 3 as shown later [in **Table 4.16**].

Since little detail will be normally available the estimate will usually be based on data from similsr previous projects updated for time, location, changes in market conditions, current design requirements and relative capacity. At this early stage the proposed project will be analyzed in fundamental elements. Among these, Pavement quantity is one used as an input variables that can be easily identified during early project schematic.

As a result, a regression model describing the total project cost as a function of pavement quantity is developed. The Regression statistics results are shown [**in Table 4.7**]. The model is useful in estimating project cost at early stages of the project since the information needed is only the pavement quantity, and so the estimation could be achieved within minutes. The developed model is

Table 4.7: Multiple regression models among total project cost and pavement quantity

Regression statistics			**Variables**	**Coef.**	**P-value**
			intercepts	**$3.46*10^{8}$**	**0.017**
R Square	**0.54**		**Asphalt work (m^2)**	**656.27**	**0.002**
Adjusted R Square	**0.51**				
Observations	**16**				
F	**15.15**				

Total project cost = 656.27X_{3} + 3.46 X 10^{8}

Where, X_{3} is the project's pavement quantity (m^{2})

It can be seen that the model has a large constant (346000000) and so in small or medium projects, the estimated total cost using the model will be not realistic e.g. for a road width of 10 m and road length of 1000m or 1km road with 7000m^{2} pavement quantity, the estimated cost is 350593890. It may give a realistic value when the intercept is not included in the model.

Table 4.8: multiple regression results among total project cost and pavement quantity without intercept.

Regression statistics			**Variables**	**Coef.**	**P-value**
			Asphalt work (m^2)	**1067.57**	**0.000**
R Square	**0.91**				
Adjusted R Square	**0.90**				
Observations	**16**				
F	**133.26**				

Total project cost = 1067.57X_3... (Model 3)

A regression model that describes the total cost of a road construction project as a function of road width and road length is developed. The coefficient of determination r^2 for the developed equation is 0.91. The regression statistics results for the developed model are shown [**in Table 4.9**]. The developed model is:

Table 4.9 multiple regression models among total project cost and road width and length

Regression statistics			**Variables**	**Coef.**	**P-value**
			intercept	7.872E8	0.186
R Square	0.36		Road width(m)	-3.635E7	0.37
Adjusted R Square	0.25		Road length (m)	4092.958	0.106
Observations	16				
F	3.32				

It can be seen that the model developed by including all variables and intercept is not at significant level of 0.05. This estimation may not give a realistic estimated value when these variables are included. Therefore, a trail was performed to develop a model without intercept value first, the results shown [in **table 4.10**].

Table 4.10: multiple regression models among total project cost and road width and length without intercept

Regression statistics			Variables	Coef.	P-value
			Road width(m)	1.437E7	0.37
R Square	0.90		Road length (m)	6478.280	0.002
Adjusted R Square	0.89				
Observations	16				
F	59.87				

Table 4.11: multiple regression models among total project cost and road length

Regression statistics			Variables	Coef.	P-value
			Intercept	3.037E8	0.174
R Square	0.31		Road length (m)	5047.835	0.031
Adjusted R Square	0.26				
Observations	16				
F	5.8				

However, it can also be seen in the model that the p-value of the road width variable is 0.37 which is higher than 0.05, meaning it is not significant to be included in the model. In order, to give a realistic estimated value only when road length is included in the model. Therefore, a trail was performed to develop a model without intercept included but, the p-value of the intercept is 0.17 which is higher than 0.05, meaning it is not significant to be included in the model. As a result, uses road length as the only input variable is formulated. The regression statistics results for the developed model are shown **[in Table 4.13**]. The formulated equation is:

Table 4.12: Multiple regression result among total project cost and road length when intercept = 0.

Regression statistics			Variables	Coef.	P-value
			Road length (m)	7888.250	0.000
R Square	0.90				
Adjusted R Square	0.89				
Observations	16				
F	59.87				

Total project cost = $7888.25X_5$ ………………………………………………. (Model 4)

Table 4.13: multiple regression result among total project cost and road size.

Regression statistics			Variables	Coef.	P-value
			Intercept	0	# N/A
R Square	0.87		Road size	747.853	0.000
Adjusted R Square	0.86				
Observations	16				
F	91.72				

A model using variables interaction has also been developed, the resulting equation is:

Total project cost = $747.85X_4$ X X5 .... (Model 5)

The coefficient of determination r^2 for the developed equation = 0.87. The regression statistics results for the developed model are shown in (**Table 4.13**).

Equations that describe the cost of the major road construction activities as a function of road length and width are developed. The best fit models are achieved when variables interaction is used and intercept value = 0. The results are shown in Table 4.14. The table shows that r^2values are high. This indicates a good relationships between dependent and independent variables.

Table 4.14: multiple regression result among major quantities project cost and road size.

Activities	Coef.	r^2	P-value	F-test	Regression equation
Earthworks	55.76	0.40	≈ 0	9.2	$55.76X_4 *X_5$
Basecoarse works	83.42	0.82	≈ 0	62.12	$83.42X_4 *X_5$
Asphalt works	109.85	0.80	≈ 0	56.83	$109.85X_4 *X_5$
Furniture cost	23.38	0.65	≈ 0	25.82	$23.38X_4 *X_5$

Then the total project cost can be calculated by summing up the cost of construction activities (model 6).

Total project cost = Earthworks cost + Base coarse works cost + Asphalt works cost

+ Furniture cost …………………………………… (Model 6)

Where,

Earthworks cost = $55.76X_4X_5$

Base works cost = $83.42X_4X_5$

Asphalt works cost = $109.85X_4X_5$

Furniture works cost = $23.38X_4X_5$

It can be seen that model 6 is similar to model 5 (both uses variable interaction), but model 6 could be used to estimate the cost of one individual construction activity (i.e. asphalt cost, base coarse cost, earthworks cost, and furniture cost). Also it has better accuracy (**MAPE**) than model 5 as shown in (**Table 4.16)**

4.3. Testing Accuracy of the Developed Models

The mean absolute percentage error (MAPE) is used to measure the accuracy of the developed models. The following formula is used to compute the MAPE (**Lowe et al., 2006**):

MAPE = $*\sum_{i=1}^{n} |(Ai - Fi)|/Ai$

Where,

A_i is the actual value

F_i is the forecast value

n is number of fitted point

Table 4.15 shows a summary of the developed regression models in the study, 10 models are developed to estimate the total cost of road construction project; 4 of them include bid quantities as independent variables (models 1 through 4), while the other two models include road length and road width as independent variables (**Models 4 through 5**).

It should be noticed that in the very early stages the bill of quantity (**BOQ**) is not available, meaning that the models using road width and length (Models 5 and 6) are more easy and fit to be used. Later, when the BOQ is available, the models based on BOQ (models 1 through 4) may be used.

Table 4.15 summary of the developed regression models

Model No.	Regression models
1	Total cost (ETB)= $26.58X_1 + 119.4X_2 + 97.62X_3$
2	Total project cost =$45.7 X_1 + 151.4X_2 +195.24X_3$
3	Total project cost = $1067.57X_3$
4	Total project cost = $7888.25X_5$
5	Total project cost = $747.85X_4 *X_5$
6	Total project cost = Earthworks cost + Sub base and Base coarse works cost + Asphalt works cost + Furniture cost Where; Earthworks cost = $55.76X_4X_5$, Sub base and Base works cost = $83.42X_4X_5$, Asphalt works cost = $109.85X_4X_5$, Furniture works cost = $23.38X_4X_5$

Where,

X1 = Earthwork; cut, fill, and topping quantities (m^3)

X2 = Sub base and Base coarse quantity (m^3)

X3 = Asphalt quantity (m^2)

X4 = Road width (m)

X5 = Road length (m)

4.4. Multiple Regression Diagnostics

An important part of regression analysis is checking that the required assumptions are met. Residual analysis was performed to evaluate the assumptions of linearity, normality, and homoscedasticity. The residuals are the differences between the observed and the predicted values. Homoscedasticity is a description of data for which the variance of the error terms appears constant over the range of values of an independent variable (inference). First, a histogram of the standardized residuals and normal probability plot was run to check the linearity and normality assumption **[18]**. Furthermore, using Q-Q plot and Shapiro normality test are used to check the normality of the error terms. The histogram, the Q-Q plot and Shapiro test results are given below.

4.4.1. RE-SAMPLING METHODS

For large samples size (n >30) statistical inference raised no question about the normality of the error term and the normality of the statistic as well. However, the study's sample size (n=15) is considered as small sample size. As Frank, H (2001) describes estimation of regression coefficients with small sample size will yield inefficient or large estimated standard error of the regression coefficients. And statistical inference with the inefficient standard error may be not acceptable or violates the normality assumption of the regression coefficient estimator. As a result the inference or the hypothesis may not be valid. This is because in the hypothesis testing about individual coefficients we use a t-statistic given by

$$t_{cal} = \frac{coefficient}{S.E}$$

Where: S.E is the standard error of the coefficient

This shows that if the S.E is large then, t_{cal} become small and the corresponding p-value become large (p>0.05). This may force to do not reject the null hypothesis that $H_o : \beta = 0$

Thus, it is necessary to obtain efficient standard error and check the distribution of the regression coefficient estimator. Hence, we use a bootstrap re-sampling method to obtain efficient standard errors and we use q-q plot and Shapiro test to check the normality of the estimated error term. The procedure of the bootstrap re-sampling method is described as follows:-

$\boldsymbol{Y = X\beta + \varepsilon}$ (the population regression equation)

$\hat{y} = X\hat{\beta}$ (the estimated regression equation using the original data)

Let us denote the estimated error as $\acute{\varepsilon} = Y - \hat{y}$ and $\hat{\beta}$ denote the estimated regression coefficients obtained from the original data.

Now to take a sample of size 15 from the original data with replacement, select 15 values from έ with replacement say (ε*) and adding the selected 15 error observations (ε*) to the original total cost (Y) observation creates a new total cost say Y* as:

$$Y^* = Y + \varepsilon^*$$

For the first iteration, the procedure consider Y* as a true total cost and then a new regression coefficients with their standard errors will be estimated and recorded. The in the next iteration it continue to select another 15 error observation with replacement from έ and create new total cost observation and then a new regression coefficients with their standard errors will be estimated and recorded. Usually 1000 or 2000 iteration are used; that means for 2000 iteration one obtain 2000 β's with respective standard errors. Finally the standard deviations of the 2000 β's are going to be the standard errors of the β's and the values of the 2000 β's also used to check the distribution of the coefficients.

The Q-Q plot with the results of Shapiro normality test of the error terms and, the corresponding results of bootstraps re-sampling methods for the selected sample of four developed model are given below:-

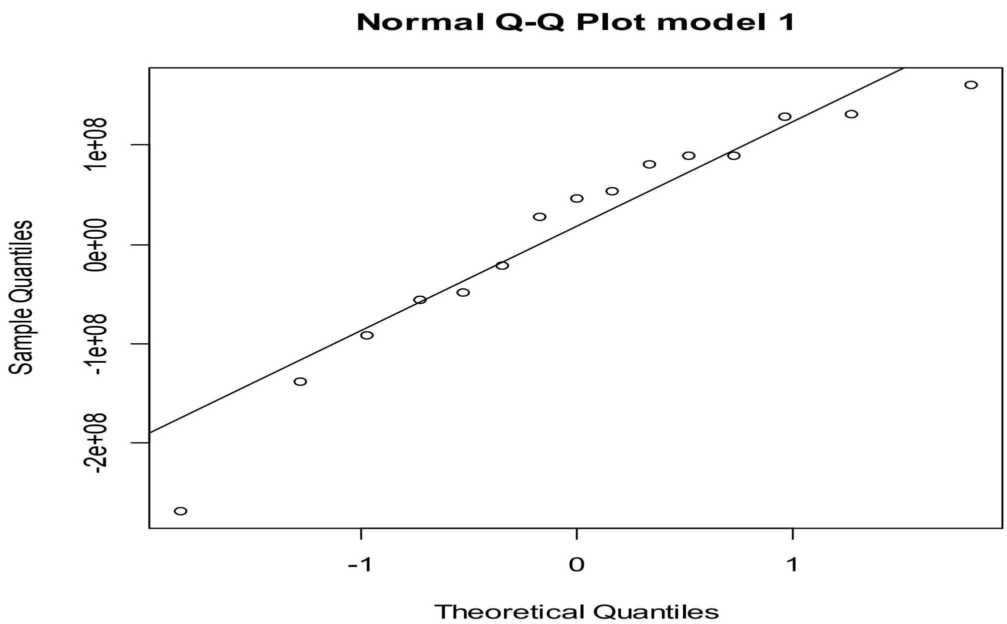


Figure 4.2: Q-Q plot for residual model 1.

As the q-q plot in figure 4.2 shows the plots are lay nearly on the straight line thus, it is acceptable to say or suggest the error terms are normal. To test the claim of the normality we use Shapiro normality test and the result is given as

Table 4.17: Shapiro-wilk normality test for the residual for model 1

	Test statistics (w)	P- value
Residual	0.928	0.2546

Thus we cannot reject the null hypothesis (H_o) that the residual follows a normal distribution. that is, models support normality assumption since p- value >0.05.

Table 4.18: Boot strap statistics result for model 1 2000 replication

Original	Bias	Standard error
89.65	1.74	24.06
684.17	7.37	111.57
448.51	4.81	88.68

The respective Histogram for each Beta is shown below

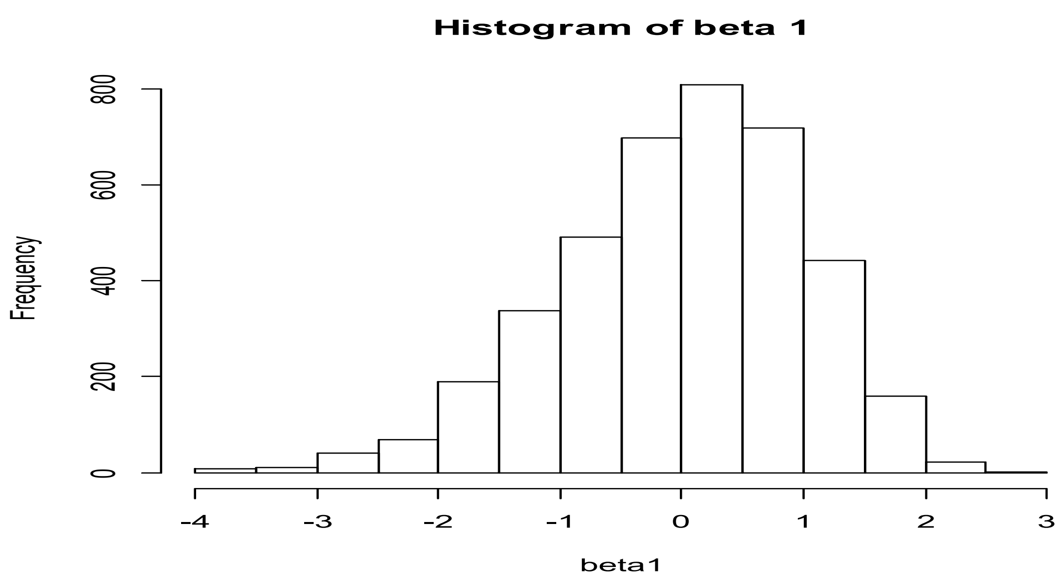


Figure 4.3: histogram for model 1 . that is(beta-mean of beta)/standard deviation of beta

Histograms developed using bootstraps re-sampling shown in the figure 4.3 in model-1 were reasonable to assume Normality. Because, from the plot one can easily suggest a normal distribution with mean 0 and variance 1.

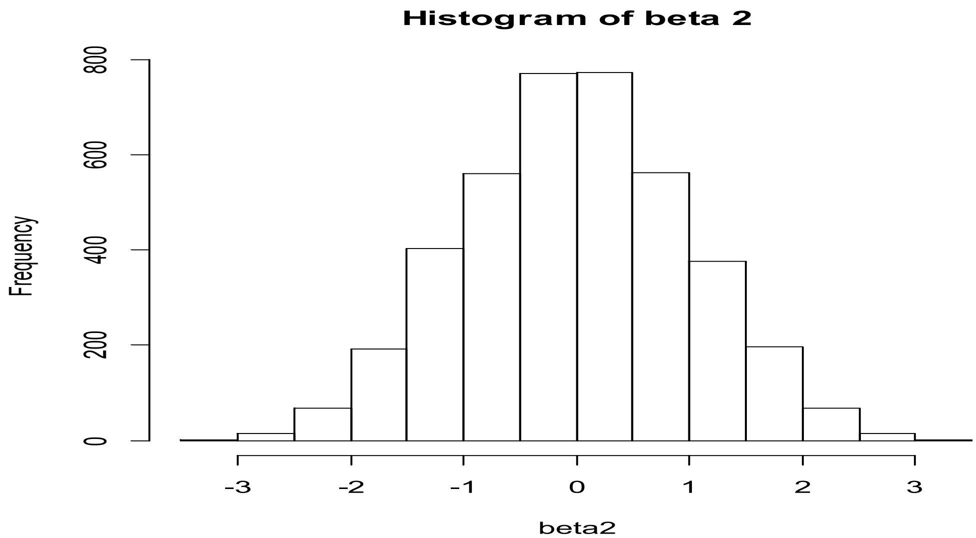


Figure 4.4: Histogram for model 1. That is (beta-mean of beta)/ standard deviation of beta

Histograms developed using bootstraps re-sampling shown in the figure 4.4 in model-1 were reasonable to assume Normality. Because, from the plot one can easily suggest a normal distribution with mean 0 and variance 1.

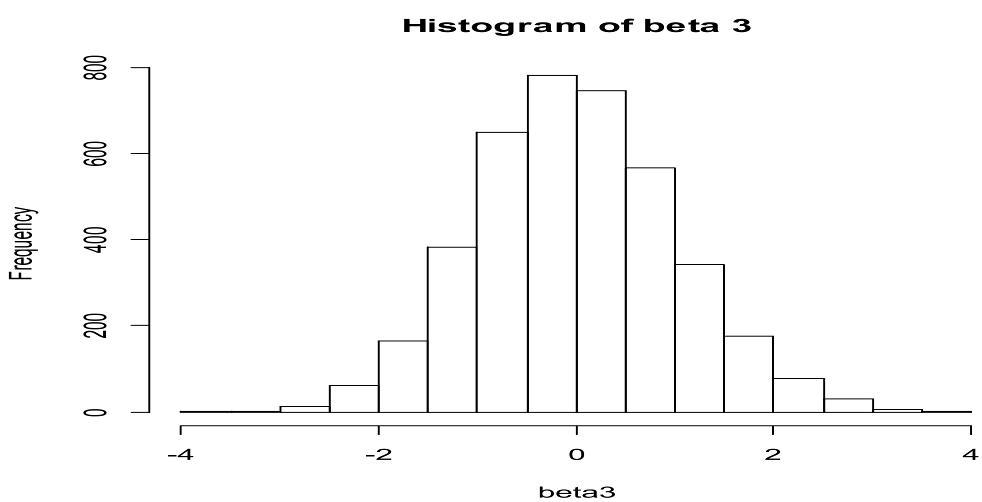


Figure 4.5: Histogram for model 1. That is(beta-mean of beta)/ standard deviation of beta

Histograms developed using bootstraps re-sampling shown in the figure 4.5 in model-2 were reasonable to assume Normality. Because, from the plot one can easily suggest a normal distribution with mean 0 and variance 1.

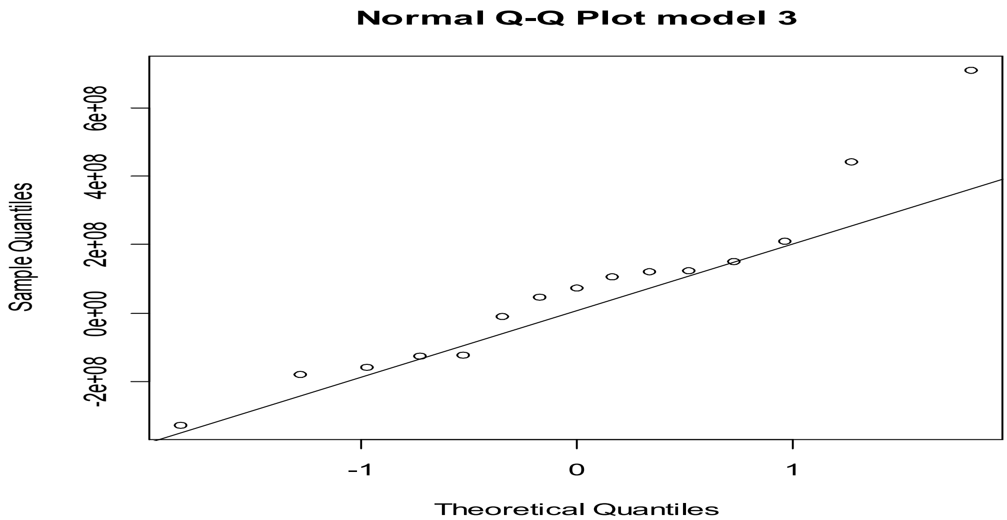


Figure 4.6: Normal Q-Q plots for residual model 3.

As the q-q plot in figure 4.6 shows the plots are lay nearly on the straight line thus, it is acceptable to say or suggest the error terms are normal. To test the claim of the normality we use Shapiro normality test and the result is given as

Table 4.19: Shapiro-wilk normality test for the residual for model 3

	Test statistics (w)	P- value
Residual	0.927	0.242

Thus we cannot reject the null hypothesis (H_o) that the residual follows a normal distribution. That is, models support normality assumption since p- value >0.05.

Table 4.20: Boot strap statistics result for model 3of 2000 replication

Original	Bias	Standard error
1867.56	12.79	86.35

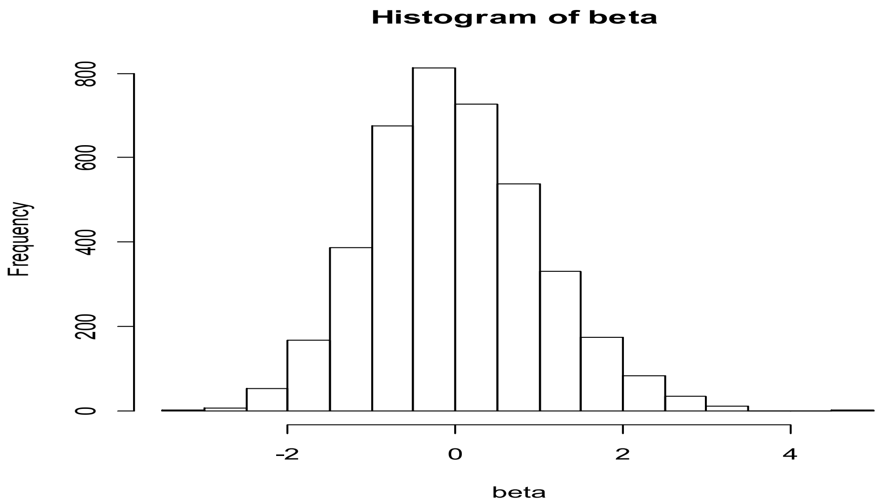


Figure 4.7: Histogram for model 3 i.e. (beta-mean of beta)/ standard deviation of beta

Histograms developed using bootstraps re-sampling shown in the figure 4.7 in model-3 were reasonable to assume Normality. Because, from the plot one can easily suggest a normal distribution with mean 0 and variance 1.

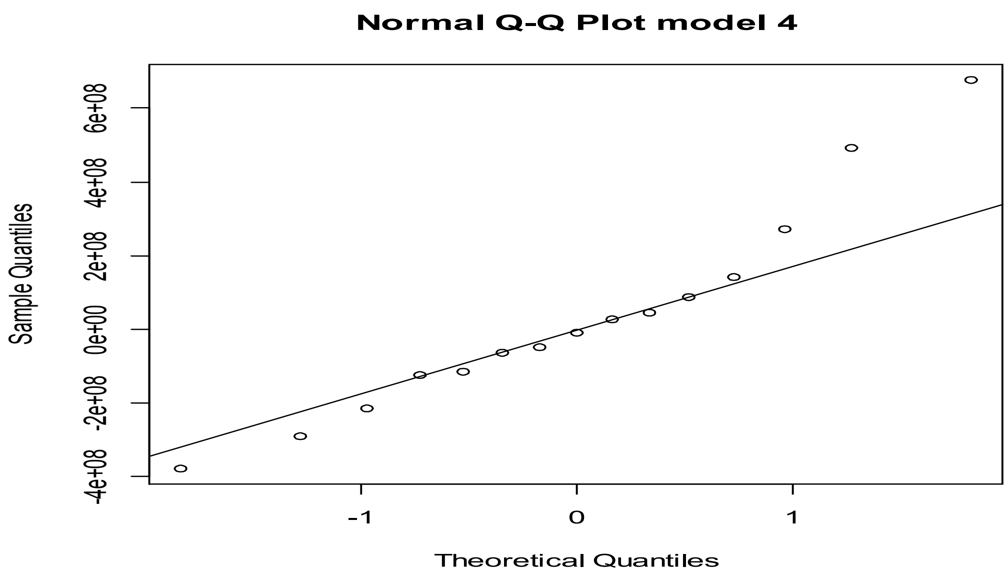


Figure 4.8: Normal Q-Q plots for model 4.

As the q-q plot in figure 4.8 shows the plots are lay nearly on the straight line thus, it is acceptable to say or suggest the error terms are normal. To test the claim of the normality we use Shapiro normality test and the result is given as

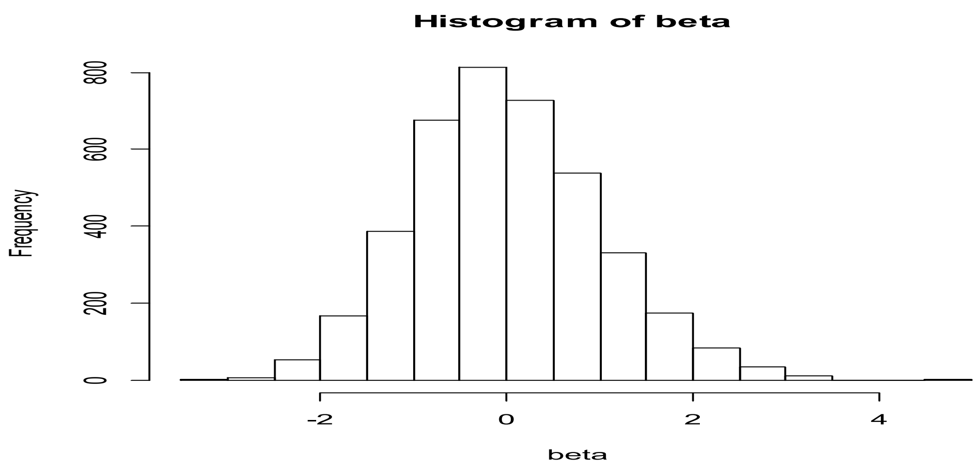
Table 4.21: Shapiro-wilk normality test for the residual for model 4

	Test statistics (w)	P- value
Residual	0.936	0.335

Thus we cannot reject the null hypothesis (H_o) that the residual follows a normal distribution. That is, models support normality assumption since p- value >0.05.

Table 4.22: Boot strap statistics result for model 4 of 2000 replication

Original	Bias	Standard error
7888.25	324.316	681.44



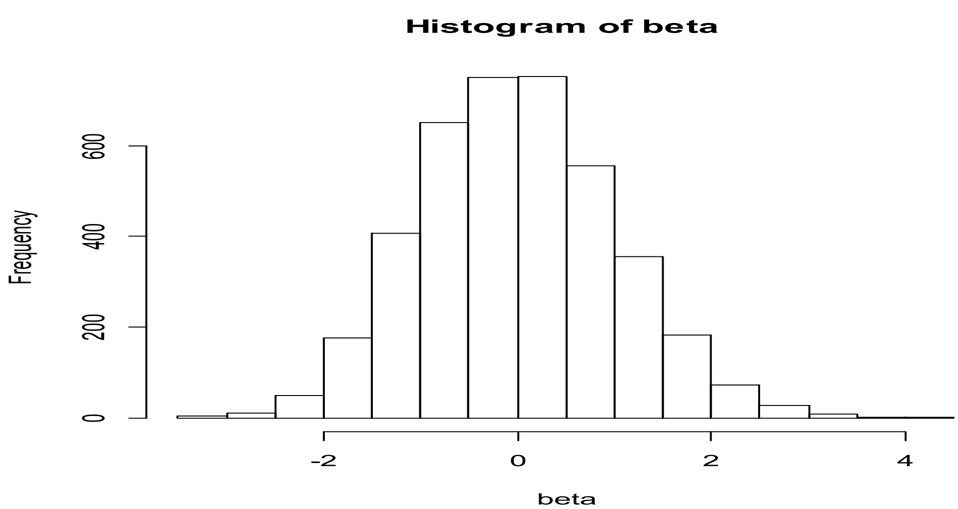


Figure 4.9: Histogram for model 4. i.e. (beta-mean of beta)/ standard deviation of beta

Histograms developed using bootstraps re-sampling shown in the figure 4.9 in model-4 were reasonable to assume Normality. Because, from the plot one can easily suggest a normal distribution with mean 0 and variance 1.

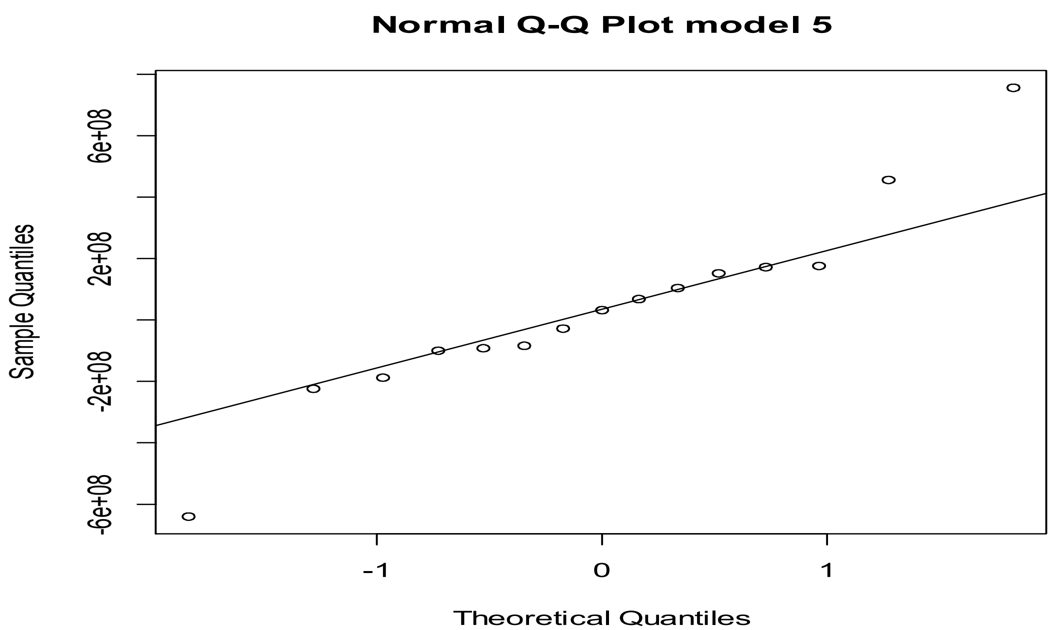


Figure 4.10: Normal Q-Q plots for model 5.

As the q-q plot in figure 4.10 shows the plots are lay nearly on the straight line thus, it is acceptable to say or suggest the error terms are normal. To test the claim of the normality we use Shapiro normality test and the result is given as

Table 4.23: Shapiro-wilk normality test for the residual for model 5

	Test statistics (w)	P- value
Residual	0.94	0.39

Thus we cannot reject the null hypothesis (H_o) that the residual follows a normal distribution. That is, models support normality assumption since p- value >0.05.

Table 4.24: Boot strap statistics result for model 5 of 2000 replication

$X_4 * X_5$	747.85	33.69	74.54

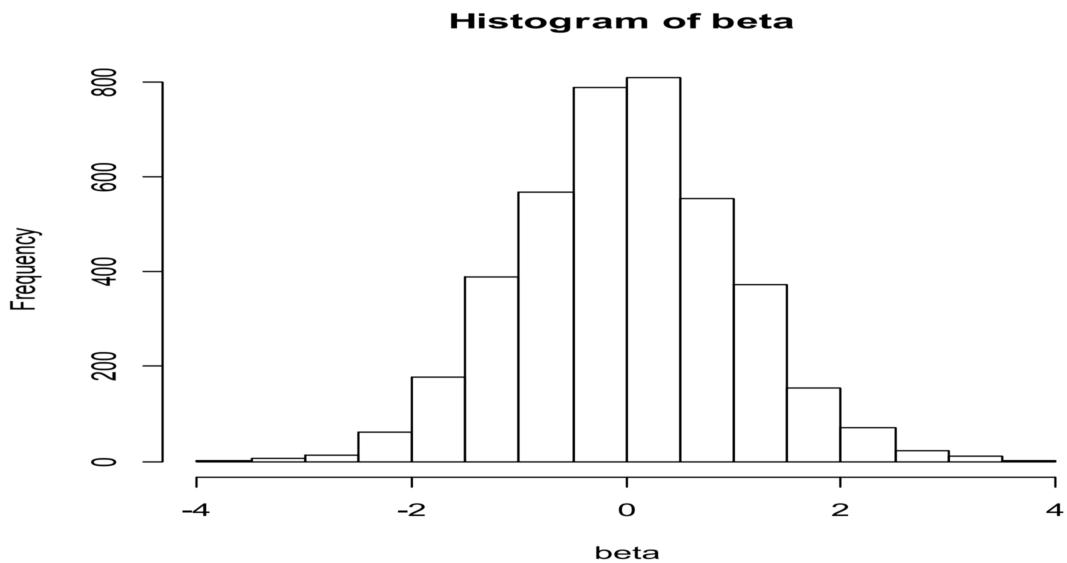


Figure 4.11: Histogram for model 5. That is (beta-mean of beta)/ standard deviation of beta

Histograms developed using bootstraps re-sampling shown in the figure 4.11 in model-5 were reasonable to assume Normality. Because, from the plot one can easily suggest a normal distribution with mean 0 and variance 1.

Overall the Q-Q plots and Shapiro normality test results of the above models shows the validity of the normal error term assumptions. And the S.E obtained using the bootstrap re-sampling method can be used to make inference or to test hypotheses. And the histograms of the coefficients obtained using the bootstrap re-sampling method shows the normality of the coefficients. So the inferences or the hypotheses testing results of this study can be considered as valid.

4.5. Results

Table 4.16 shows the **MAPE** resulting from using the developed regression models to estimate the total cost of 16 data sets of road construction projects. The results show the following:

- In line with lowe et al.,(2006); Schexnayder et al., (2005) the MAPE of the developed models is ranging from ±16.3% to ±38.9% with coefficient of determination in this study is between 87 and 98 respectively which compare favorably with past researches which have shown that the estimate accuracy in the early stages of a project ±25% for preliminary and ±50% for conceptual cost estimation repectively.
- **The table 4.16** shows that Linear model for estimating total cost road construction project can be modeled within MAPE ranging from ±16.3% to ±38.9% and the coefficients of determination for these models were between 87 to 97.
- Unlike to other researches, that uses large size samples in running regression models, in this study (small sample size) to check or diagnose the developed models we used a bootstrap re-sampling method using R-software. And the following assumptions of the model are checked:-
    - Normality of the residuals
    - Heteroskedasticity
    - Outliers
    - Collinearity
- It can be seen that the models that use bid quantities as independent variables are more accurate than those using road length and road width as independent variables
- **The table 4.16** shows that for the models that use bid quantities as independent variables, when the number of work items involved in the model increase, the r2 value increase and MAPE value decrease.
- **The table 4.16** shows that for the models that use road length and width as independent variables, the MAPE is ranging from ±32% to ±39%. It shows that the model with highest accuracy is the model includes road length only with zero intercept value (model 4), while the model with least accuracy is the model includes road length and width (**Model 5**)
- **The table 4.16** shows that except for model 2, the models that use BOQ as independent variables, the MAPE is ranging from ±17% to ±25%. It shows that the model with highest accuracy is the model includes all construction activities (**Model 1**), while the model with least accuracy is the model includes pavement quantity only (**Model 3**).

Table 4.16: Mean absolute percentage error (MAPE) and r2 of the developed regression models.

Model No.	MAPE	r^2
1	16.3%	.98
2	45%	.65
3	25.4%	.91
4	32.3%	.90
5	38.9%	.87
6	35.3%	.65

CHAPTER FIVE

5. CONCLUSION AND RECOMMENDATION

5.1. Conclusion

This research examined the reliability of estimates for road construction costs of road projects in Ethiopia, and indicates how the findings can be used for the selection and improvement of these projects. In this study in order to achieve its objective first the correlation between major activities in road construction were identified. Next, taking all consideration for model development criteria and based on regression analysis early cost estimation models were develop at a significant levels, Finally, the reliability of these models checked and evaluated with previously done research works. This can be done by comparing the mean percentage errors of models, what is already discussed under result of this study. Therefore these models can be applicable in future during project development according to the level of project cost estimating. But it is vital that implementer's or cost estimator's consider other project characteristic and factors that significantly associated to estimation of future costs.

The focus of this research has therefore been developing early cost estimating models for road construction projects using multiple regression techniques using past experience of the Federal Road projects. The models were developed based on 16 set of data collected in the Federal Road projects in Ethiopia. Such types of models are very useful, especially in its simplicity and ability to be handled by calculator or a simple computer program. It has a good benefit in estimating project cost at early stages of the project since the information needed could be extracted easily from sketches or scope definition of the projects.

It must be remembered that an estimated project cost is not an exact number, but it is opinion of probable cost. The accuracy and reliability of an estimate is totally dependent upon how well the project scope is defined and the time and effort expended in preparation of an estimates.

In this study, six regression models are developed; out of which two of them are served as preliminary cost estimating models of road construction projects as a function of quantities of road construction activities. The remaining three models are used to develop early cost estimating models of road construction projects as a function of project size (i.e. road length and road width) include road length and width. The coefficients of determination (r^2) for the developed models range from 0.87 to 0.95. Since the number of samples in the model development is so small it raises the question of reliability of the developed estimation model. To check and prove, whether the models are good and the predicted values from the forecast models fit to apply in forecasting future projects techniques called bootstrap were used.

The values of the mean absolute percentage error (**MAPE)** of the developed regression models are ranging between ±38.9%, the results compare favorably with past researches which have shown that the estimate accuracy in the early stages of a project at conceptual stage were between ±50% and for preliminary cost estimation the developed regression cost models are ranging between 25.6%, the results also compare with past research that estimate accuracy is nearly the same as the figure ranging between ±25%.

5.2. Recommendation

One common element in all the road projects has been the use of forecasting estimates of cost for the future projects, which is gaining wide acceptance in anticipating reliable estimates of projects cost. Cost estimation done during conceptual phase of the project cycle usually calculated approximately which leads to great inaccuracy. The reason for that is to have rough idea of a project in financial terms. On the other hand, project details are not fixed and well defined due to lack of information. In such cases regression models could be a suitable tool to estimating project cost.

Initially a lot of time was spent to collect road project data that better define total cost to develop and test models. However, the effort made to get these data from the client owns database or consultant's is not sufficient to develop and test models. Finally considering at the data's quality and sample size this study uses bootstrap methods of re-sampling to overcome such difficulties using R-software. Besides, regression techniques were used as a tool to develop forecasting models, which is more compatible to quantitative data sets.

The critical key issue in this study to develop cost models at this level lies in how to build a realistic cost estimates between parties based on the experience gained from this research study. In addition, it addresses mean percentage errors expected to occur during cost estimation process in accordance with level of project cost estimation stage. As a result of the research findings project stakeholder's clients, financers, and government legislator's makes use of these cost models to support their decision-making. This condition has advantage by creating communication between stake holders and professionals easier. Credit times given by project donors or financer's should truly reflect realistic estimates. Validation of the proposed model very much depends on utilization of appropriately recorded historic data.

To address the above challenges, this research would recommend project stakes holders client, government legislators, and financers to put particular emphasis or to take measures to promote this kind of research out puts by sponsoring scholars in their commitment to produce research works on early estimates of future infrastructure investments realistically, due this study can serve as a stepping stone for further research on the topic to have improved cost estimation practice and to have a better end result as follows:

- The aim of the research output in the context of the Ethiopian road sector development are:
  - i) To ensure that public infrastructure development funds are used economically.
  - ii) To ensure that a convincing estimating method can be made for public infrastructure projects to benefit from external funding.
- Budgeting of projects using forecasting models should be conducted when it can be justified by
  - i) The size of the project
  - ii) The likelihood of a significant similarity between current market, location, and technology.
- It is likely that the application of forecasting costs may be difficult in the first instant because it will require a more detailed specification of scope than sometimes provided in project reports, location indexes, and forecast cost index for the intended period of projects. The usefulness of the application of forecasting costs depends critically on the reliability of estimates of quantities made at project feasibility stage. It is important to improve these aspects of project design and analysis if the full values of using forecast costs are to be realized.
- It is clear that the values of forecast model costs are likely to be changed over time with changes in economic policies and other conditions such as demand and supply of materials, equipments, and labor national wise or globally. It is also likely that some of the estimates will require correction in the light of future trends or knowledge. It is therefore strongly recommended that forecast cost or price estimates are reviewed periodically.
- It guides decision of feasibility of producing the within the authorized figures and provides a means of subsequent checking and control of overall expenditures realistically, usually in the form of parametric techniques introduced in this research.
- It permits regular reassessment of the overall cost estimating results and point out changes resulting from revised project requirements and specification in order to use recent and compatible cost estimation methods.
- For the project sponsoring organization, accurate cost estimates are vital for decisions that include strategies for potential project screening, and resource commitment for further project

development. Therefore they insist or enforce their counterparts on improving historic project cost data records in consistent and regular manner.

- Appropriate parametric cost models should be developed to all stage of project level to help solve problems which come from a number of uncertainties such as cost estimation at the conceptual phase.
- A comprehensive and reliable cost database should be established with the regional integration and trade division by joint effort of all stakeholders of the construction industry. The database should be updated and recorded regularly at short time interval to give realistic price data. An example for minimum and required, highly recommended data set by road cost knowledge system (ROCKS) where given in the appendix E of this research study.

5.2.1. Recommendations for Further Works

This research has identified statistically significant explanatory variables that affect early total road cost and also develop a total road cost model on the basis of the variables using regression analysis. The research moreover, has forwarded (qualitative) theoretical interventions and recommendations that can improve estimation of early total cost. Because of lack of information, in the model development, this study do not incorporate qualitative variables that can able to affect the estimation of early total cost such as, experience of the expert, the nature of the location and others. Thus to improve early total cost estimation it will be better if the qualitative variables are included in model development with appropriate statistical methods.

References

[1] Albert Lester (2006) Project Management, Planning and Control: Managing Engineering, Construction and Manufacturing Projects to PMI, APM and BSI Standards: Elsevier Science & Technology Books, 5th Edition.

[2] Ahuja, Hira N., Dozzi, S.P., and Abourizk, S.M. (1994). "Project Management: Techniques in Planning and Controlling Construction Projects", John Wiley & Sons, Inc., NY.

[3] Ahuja, H.N. and W.J. Campbell, Estimating: From Concept to Completion, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1987.

[4] Allison S., Allison D. Multiple Regression: A Primer. Forge Press. 1999

[5] Becker H. J. and Behailu D.D. (2006). Public Private Partnership in Road Projects in Ethiopia.

[6] Bell, L. C. and Bozai, G. A. (1987). Preliminary cost estimating for highway construction projects, AACE Transactions, C.6.1–C.6.4

[7] Bent Flyvbjerg, Mette Skamris Holm, and Søren Buhl, (2002). Journal of the American Planning Association, Vol. 68, No. 3,American Planning Association, Chicago, IL.

[8] Black, J. (1984) Application of parametric estimating to cost engineering, AACE Transactions, B.10.1–B.10.5.

[9] Brikes D, Birkes D, Dodge Y. Alternative Methods of Regression. Wiley-Interscience. 1993

[10] C. Hendrickson (1998) Project Management for Construction: Fundamental Concepts for Owners, Engineers, Architects and Builders, Pittsburgh: Prentice Hall.

[11] Construction Industry of Ethiopia December (2013). Construction industry of Ethiopia Amharic Version, pp 10-15

[12] Christian, J. and Newton, L. (1998) Highway construction and maintenance costs. Construction Engineering and Management Group, University of New Brunswick, Fredericton, NB E3B 5A3, Canada

[13] Ethiopian Road Authority march (2010). RSDP performance report retrieved from Ethiopian Road Authority website http//www.era.gov.et//performance

[14] Flyvbjerg B. H., Mette S., and Buhl S. Underestimating costs in public works projects, error or lie, Journal of the American Planning Association, Vol. 68, No.3, pp 279-292, 2002

[15] Frank H. Regression Modeling Strategies: with Applications to Linear Models, Logistic Regression and Survival Analysis. Springer. 2001

[16] Getachew Tsegaye (2009). Design Risk Management In Ethiopian Federal Road Projects

[17] Gould, F. (2005). Managing the construction process: estimating, scheduling, and project control, Pearson Education, Inc., New Jersey

[18] Hair, Joseph F., Anderson, Rolph E., Tatham, Ronald L., and Black, William C. Multivariate Data Analysis. 5th Edition, Prentice-Hall, NJ, 1998.

[19] Han, S. H., Kim, D. Y., and Kim, H., (2008) Two-staged Early Cost Estimation for Highway Construction Projects, Report (2008), Civil & Environmental Engineering, Yonsei University, Seoul, Korea

[20] Hegazy, T. and Ayed, A. (1998) 'Neural network model for parametric cost estimation of highway projects', Journal of Construction Engineering and Management, 124 (3), 210-218

[21] Hendrickson, Chris and Au, Tung (1989) Project Management for Construction: Fundamental Concepts for Owners, Engineers, Architects, and Builders, 1st edition, Prentice Hall, London

[22] Hojjat Adeli and Asim Karim, (2005) Construction Scheduling, Cost Optimization, And Management

[23] Holm, L., Schaufelberg, J., Griffin, D., and Cole. T. (2005) Construction cost estimation processs and practices, Pearson education, Inc., New Jersey

[24] Jelen, F. and Black, J. (1983)Cost and Optimization engineering,2nd edition, McGraw-Hill, New York

[25] James P. Lewis (2007) fundamentals of project management. USA: Ameican Management Association, 3rd Edition.

[26] Jameson, A. and Nugent, H. (2006). 'Delays ahead as road costs soar', The Times, Monday, 23 January 2006, p. 3

[27] Kaplan R. S (2010). Conceptual Foundations of the Balanced Scorecard, Boston: Harvard Business School Press.

[28] Lewis-Beck M. Applied Regression: An Introduction. Sage Publication Inc. 1980.

[29] Little, R. J. A. 1988. A test of missing completely at random for multivariate data with missing values. Journal of the American Statistical Association, 83, 1198-1202.

[30] Martin, J. (2007). 'It's always a forecast', Journal of RICS Construction, February, pp. 16–17

[31] Martin Skitmore; Factors Affecting Accuracy of Engineers' Estimates; Transactions AACE 1988

[32] Merrow, E. W. (1988). Understanding the outcomes of megaprojects: A quantitative analysis of very large civilian projects. Santa Monica, CA: RAND Corporation.

[33] Miles J., Shevlin M. Applying Regression and Correlation: A Guide for Students and Researchers. Sage Publication Inc. 2001.

[34] Martínez-Zarzoso and Nowak-Lehmann (2006). Is Distance a Good Proxy for Transport Costs? The Case of Competing Transport Modes, CeGE-Discussion Paper 54, May

[35] MoFED National Accounts Department, National Accounts statistics of Ethiopia. Nov. 2006

[36] FDRE Central Statistical Agency 06 November 2013

[37] Nijkamp and Ubbels (1999) National Audit Office/Department of Transport 1992; World Bank 1994;

[38] Norman, A. (1994) 'Cost Estimating', notes presented at the UMIST Project Management Course, UMIST

[39] PMBoK (2013) project Cost management to Project Management Body of Knowledge:, Project Management Institute USA.

[40] Pearce A. R., Gregory, Rita A., and Williams L. (1996) "Range Estimating for risk management using artificial neural networks", Journal of Parametric Estimating, 14(1), pp 88-93.

[41] Rose, A. (1982) 'An organized approach to parametric estimating', Transactions of the seventh International Cost Engineering Congress. London, England

[42] ROCKS, Roads Costs Knowledge system (v.2.2), (The World Bank, Transport Unit-TUDTR), 2004. http://www.worldbank.org/transport/roads/rd_tools/rocks_main.htm. Assessed march10,2013

[43] Sanders, S.R.; Maxwell, R.R.; and Glagola, C.R. (1992) 'Preliminary estimating models for infrastructure projects', Cost Engineering, 34 (8)

[44] Schexnayder, C. J. and Mayo, Richard E. (2003) Construction Management Fundamentals, McGraw-Hill Higher Education, Boston, MA

[45] Sodikov, Jamshid (2005) 'Cost estimation of highway projects in developing country: Artificial neural network approach', Journal of the Eastern Asia Society for Transportation Studies, 6, 1036 – 1047.

[46] Schexnayder, C. J. and Mayo, Richard E. (2003) Construction Management Fundamentals, McGraw-Hill Higher Education, Boston, MA

[47] SPSS (1997), Missing Value Analysis 16.5, MaryAnn Hill / SPSS Inc.

[48] Tsunokawa, K. T. (1983) "Evaluation and Improvement of Road Construction Cost Models." Unpublished draft working paper prepared for the World Bank. Washington, D.C. World Bank.

[49] W.F.CHEN, J.Y.RICHARD LIEW 2003 The Civil Engineering Hand book 2 edition page, 25

[50] W.F. chen, J.Y. Richard Liew (2003) Civil engineering handbook 2 edition page 25

[51] Wideman, R. M., (1995) Cost Control of Capital Projects, BiTech Publishers Ltd, Richmond, BC, Canada.

[52] Washington State Department of Transportation., (2014) Cost Estimate process retrieved from the Washington State Department of Transportation website http//Www.wsdot.wa.gov//process

APPENDIX

Appendix A: Ethiopian Roads Network Map

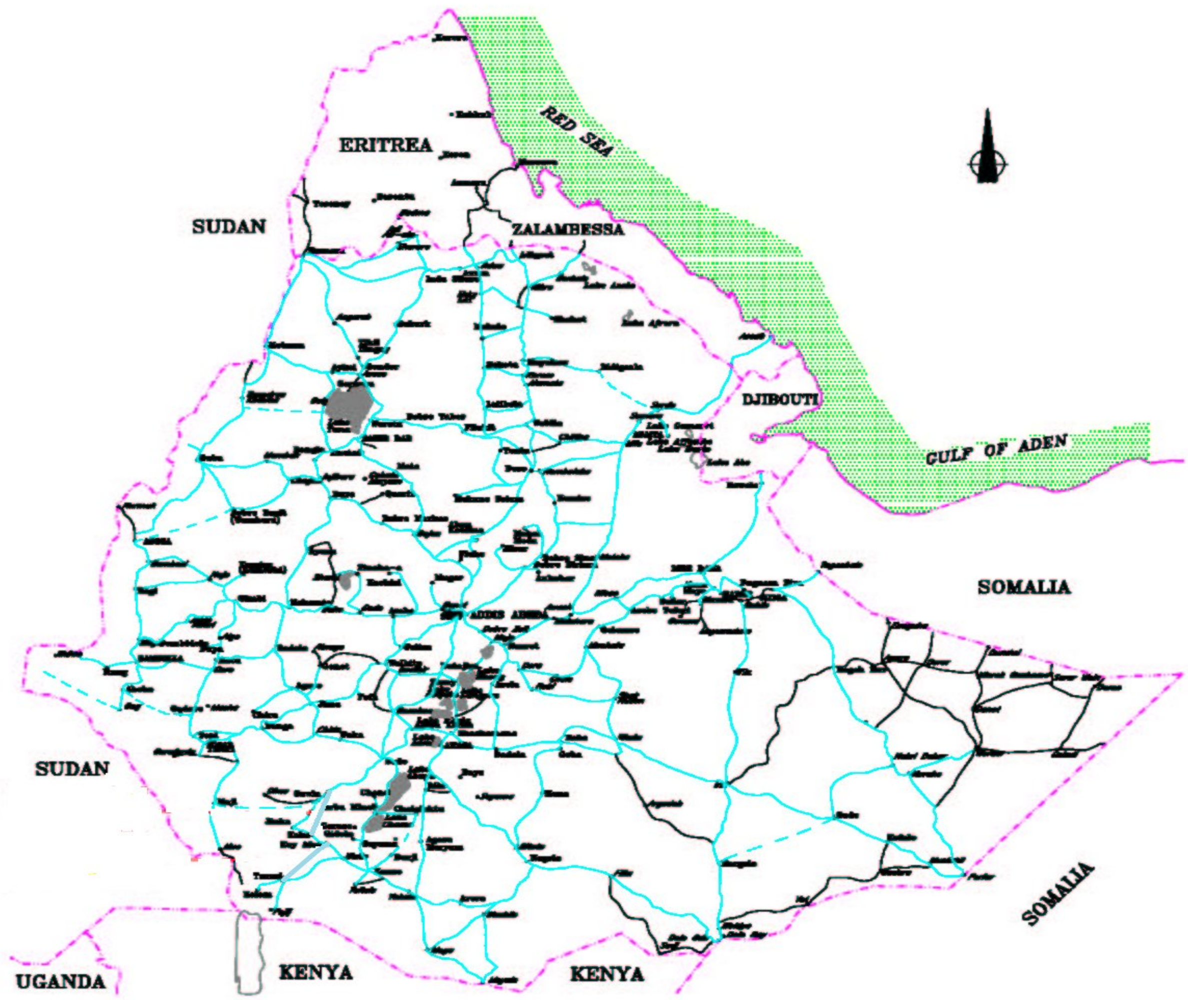


Figure A1.Existing Roads Network in Ethiopia (Source: ERA ERDP Project Status Report 2010)

Appendix B: Data Collection and Processing

The first task of the research reported in this paper was to establish a large size sample of Road construction projects however we collect smaller size than what is common in this area of research, a sample small that is not enough to allow statistical analyses of costs. Here a first problem was that data on actual costs in transportation infrastructure projects are relatively difficult to come by. One reason is that it is quite time consuming to produce such data. For public sector projects, funding and accounting procedures are typically unfit for keeping track of the multiple and complex changes that occur in total project costs over time. For large projects, the relevant time frame may cover four, six, or more fiscal years from decision to build, until construction starts, until the project is completed and operations begin. Reconstructing the actual total costs of a public project, therefore, typically entails long and difficult archival work and complex accounting. Unfortunately, this also tends to keep data from the hands of scholars. And for both public and private projects, data on actual costs may be held back by project owners because more often than not, actual costs reveal substantial cost escalation, and cost escalation is normally considered somewhat of an embarrassment to promoters and owners. In sum, establishing reliable data on actual costs for even a single transportation infrastructure project is often highly time consuming or simply impossible.

This state of affairs explains why small-sample studies in this field of research. But despite the problems mentioned, after half-years of data collection and refinement, we were able to establish a sample of forty road construction projects with data on both actual construction costs and estimated costs at the time of decision to build. The project types are new construction, gravel road construction, RR50, and Asphalt concrete. The projects are located in thirteen regional states of the countries within five regional administrative offices in the main office of Ethiopian Road Authority.

The projects were completed between 1997 and 2013. Older projects were included in the sample in order to test whether the accuracy of estimated costs improved over time.

In statistical analysis, data should be a sample from a larger population, and the sample should represent the population properly. These requirements are ideally satisfied by drawing the sample by randomized lot. Randomization ensures with high probability that factors that cannot be controlled are equalized. A sample should also be designed such that the representation of subgroups corresponds to their occurrence and importance in the population. In studies of human affairs, however, where controlled laboratory experiments often cannot be conducted, it is frequently impossible to meet these ideal conditions.

We select the projects for the sample on the basis of data availability. All projects that we knew of for which data on construction cost development were obtainable were considered for inclusion in the sample.

Cost development is defined as the difference between actual and estimated costs in percentage of estimated costs, with all costs measured in fixed prices. Actual costs are defined as real, accounted costs determined at the time of completing a project. Estimated costs are defined as budgeted, or forecasted, costs at the time of decision to build.

Data on cost development were available for forty-eight projects. We then rejected twenty-four projects because of insufficient data quality. For instance, for some projects we could not obtain a clear answer regarding what was included in major variables in cost estimation models and similarity of works and specification. More specifically, of those sixteen projects, we rejected one because we could not establish whether or not cost data were valid and reliable. We rejected nigh teen projects because they had been completed before 2008 and no bid quantity is obtained for major road activities. Finally, eight projects excluded because cost development for them turned out to have been calculated before construction was completed and operations begun; therefore, the actual final costs for these projects different from the cost estimates used to calculate cost development, and no information was available on actual final costs.

Appendix C: All Project Data Set for Correlation Evaluation

Project No.	Project	Cc*	Tc	Cf*	Tf	L(km)
1	Addis - Modjo - Awasa	824569353	1096	1002136689	1487	263400
2	Modjo - Awash Arba	584290589	1096	568173038	2102	160300
3	Awash Arba - Gewane	493230380	912	580878000	1196	135800
4	Gewane - Mille	555787324	912	899753495	1218	146030
5	Logia - Semera (Contract 1)	202988189	720	226575216	1080	41000
6	Tarma Ber - Kombolcha+Check Variation	1127611294	900	960968477	2294	187000
7	Addis - Gohatsion Section 1 (5.5 - 31.5 km)	145787445	522	133863239	522	26000
8	Addis - Gohatsion Section 2 (31.5 - 76.5 km)	145787445	522	133863239	522	45000
9	Addis - Gohatsion Section 3 (76.5 - 95.0 km)	145787444	871	128820558	871	18500
10	Debre Markos - Merawi	1026472149	1096	859205933	1906	219800
11	Merawi - Gondar	584290589	1096	568173038	2102	208000
12	Awash - Hirna	650878953	1096	765896063	1887	140400
13	Hirna - Kulubi	483167684	1203	571394840	2382	91000
14	Gob Gob - Gashena (AC)	358719178	1278	615673242	1603	86000
15	Gashena-Woldiya Cont.3 Woreta-Woldiya (AC)	699603410	1278	936731233	1971	107700
16	Azezo-Metema Lot-1 (AC)	429619427	1095	580322497	1214	87000
17	Azezo-Metema Lot-2 (AC)	432653690	1095	536192098	1214	98000
18	Gondar Debark Road Upgrading Pr. (AC)	690779965	1215	1033100608	1520	100000
19	Shire-Shiraro Humera Lugdi Lot.2 (AC)	878792804	1095	1146968702	2190	161000
20	Nazreth - Assela Contract-3 DODOLA JUNCTION-GOBA IPC-53 (AC)	694923439	2207	1319068312	4360	116000
21	Harar-Jijiga (AC)	879183739	1095	768192809	2291	103600
22	Dembi -Bedele (AC)	352296736	730	275054025	2190	62000
23	Mojo Ejere Arerti Contract-2 ARERTI-GOBENSA (AC)	444696050	913	555164650	2756	36000
24	Degehabur - Kebridehar Cont.1 Degehabur-shekosh	422343483	1096	482768415	1874	90000

	(AC)					
25	Shire - Shiraro - Humera Lugdi lot.1(AC)	616441995	1095	1107347609	2190	156000
26	Nazareth Assela Dodola & Shashemene Goba Cont. 2 Assela Dodola Junc. ipc dodola Hard copy(AC)	248273351	1095	1079715789	2190	79000
27	Addis - Tarmaber (Contract 1)Rehabilitation5/25/2007(AC)	1214429190	1095	1228048146	2190	70000
28	Addis Tarmaber(contract 2)Rehabilitation5/30/2007(AC)	783149497	1095	792606060	1825	115000
29	Wereta - Gob - Gob Lot-2 (DBST)	401092586	1095	439810305	1491	49900
30	Kulubi - Dengego - Dengeno - Dire Dawa (DBST)	411466527	913	526760940	1557	80000
31	Woldiya - Alemata (DBST)	386178790	1096	463348281	2425	78300
32	B/Mariam - Wukero (DBST)	522536021	1096	475057724	1987	273260
33	Mekenajo - Nejo (DBST)	311188811	720	263824045	1018	61000
34	Nejo - Medi (DBST)	331770149	720	299405508	896	70000
35	Alemgena - Butajira (DBST)	550863797	1472	540314286	1247	118000
36	Butajira - Husana (DBST)	491088565	912	303310942	1028	100000
37	Assosa/Sherkole – Blue Nile Road Project, IPC No 25 (RR50)	280038383	1095	382541125	1125	70400
38	Assosa/Sherkole – Guba Road Project (RR50)	328181079	1095	421755683	1447	57112
39	Tongo Begi Mugi Road pr.cont 1 Tongo Gidami (Gravel)	136120676	1234	152724498	2094	70000
40	Ilya - Adura (Gravel)	335013491	1095	241937341	2521	88400

		Adjusted contract Amount vs. adjusted final amount
ALL	Correlation (Cc* - Cf*)	0.778063366
AC	Correlation (Cc* - Cf*)	0.743039887

*Denotes adjusted contract or final amount

Appendix D: AC Projects for Running Regression Analysis

Project	*ATPC	BOQ EW	BOQ BC	BOQ AW	year of compl	year of start	KM
Gob Gob - Gashena (AC)	$6.2x10^8$	823000	375600	743114	10-Jan-09	4-Jul-06	86
Gashena-Woldiya Cont.3 Woreta-Woldiya (AC)	$9.4x10^8$	519500	462500	99200	1-Dec-11	27-Jul-06	108.55
Azezo-Metema Lot-1 (AC)	$8.3x10^8$	661898.8	561800	657000	27-Dec-08	27-Jul-06	87
Azezo-Metema Lot-2 (AC)	$7.6x10^8$	976000	396200	725000	20-Jul-09	27-Jul-06	98
Gondar Debark Road Upgrading Pr. (AC)	$1.3x10^9$	1595000	698505	785000	3-Feb-12	5-Dec-08	99.2
Shire-Shiraro Humera Lugdi Lot.2 (AC)	$9.8x10^8$	2370639	891664	1223549	10-Apr-10	10-Apr-07	161
Nazreth - Assela Contract-3 DODOLA JUNCTION-GOBA IPC-53 (AC)	$9.4x10^8$	800320	580775	1067933	21-Sep-07	10-Aug-04	116
Harar-Jijiga (AC)	$7.7x10^8$	927043	534589	836989	15-Dec-08	16-Dec-05	103.6
Dembi -Bedele (AC)	$2.8x10^8$	272862	444740	189048	4-Jan-08	4-Jan-06	62
Mojo Ejere Arerti Contract-2 ARERTI-GOBENSA (AC)	$5.6x10^8$	4030835	141400	420300	21-Nov-08	22-May-08	36
Shire - Shiraro - Humera Lugdi lot.1(AC)	$1.1x10^9$	1180366	454749	624010	5-Dec-12	5-Dec-08	156
Nazareth Assela Dodola & Shashemene Goba Cont. 2 Assela Dodola Junc. ipc dodola Hard copy (AC)	$1.1x10^9$	474000	48400	91000	27-Jul-12	27-Jul-07	79
Addis - Tarmaber (Contract 1)Rehabilitation5/25/2007(AC)	$1.2x10^9$	2541871	758557	953382	12-Jun-12	25-May-07	70
Addis Tarmaber(contract 2)Rehabilitation5/30/2007(AC)	$7.9x10^8$	793663.5	408975	627683	30-May-12	30-May-07	115
Wereta - Gob - Gob Lot-2 (DBST)	$4.4x10^8$	394060	282538	369024	5-Feb-09	5-Apr-06	49.9

*TPC: Adjusted total project cost

Appendix E: Data List Domestic Contractors

Project No.	Project	Status	Contract Signed	Completion Date	Duration (Days)	Contract Duration	Contract Amount	Final Amount	Length (km)
1	Diko – Chira	C	22 Jun. 99	5 May. 2000	683.00	455.00	10,477,539.33	10,477,539.33	18.70
2	Dubti - Asayita (DBST)	C	5 June. 2001	27 June. 2001	1,694.00	730.00	105,584,730.00	89,687,337.21	57.82
3	Alemata - Korem - B/Mariam (Own - AC)	SC	1 Dec. 00	1 March. 07	2,250.00	2,130.00	21,575,000.00	397,388,778.94	68.00
4	Alemata - Muhane – Maychew	C	10 Feb. 99	1 May. 2000	446.00	365.00	68,880,600.00	87,576,104.04	67.40
5	Shire - Shiraro - Humera I	C	5 Feb. 94	22 Aug. 99	2,022.00	730.00	18,217,309.83	30,276,030.75	70.50
6	Shire - Shiraro - Humera II	C	5 May. 94	23 Aug. 97	1,292.00	730.00	39,591,367.14	41,146,137.15	69.08
7	Humera - Ludgi	SC	16 Dec. 04	6 April. 07	845.00	455.00	57,663,337.21	57,169,230.63	67.00
8	Tekeze - Humera	C	9 June. 97	05 April. 03	2,126.00	1,092.00	154,271,535.40	176,109,267.89	127.77
9	Ghint - Wohni	C	5 June. 01	576 c days	576.00	365.00	33,573,414.89	30,281,715.50	39.00
10	Mohane - Hiwane	C	27 June. 00	571 cdays	571.00	644.00	128,348,934.77	125,237,159.79	67.78
11	Wohni - Shehedi	C	27 June. 00	733 cdays	733.00	365.00	24,429,668.94	26,273,574.70	25.50
12	Shehedi - Kokit	C	27 June. 00	798 days	798.00	538.00	24,429,668.94	26,273,574.70	69.80
13	Kokit - Metema	C	27 June. 00	800 cdays	800.00	365.00	25,285,823.38	30,734,594.07	23.98
14	Fincha - Lemlembereha	SC	12 Sep. 02	12 Sep. 04	1,673.00	720.00	74,926,964.87	67,315,332.38	43.60

	Check								
15	Shkhusien - Mecheta II	SC	16 Aug. 02	1 April. 07	1,692.00	1,440.00	124,728,864.00	116,338,160.73	145.00

Appendix F: Data List Foreign Contractors

Project No.	Project	Status	Contract Signed	Completion Date	Duration (Days)	Contract Duration	Contract Amount	Final Amount	Length (km)
1	Addis - Modjo – Awasa	C	12 Sep. 97	9 Nov. 2001	1,487.00	1,096.00	310,979,872.92	386,197,700.31	263.40
2	Modjo - Awash Arba	C	19 March. 99	31 Dec. 04	2,102.00	1,096.00	227,449,546.00	254,539,566.00	160.30
3	Awash Arba – Gewane	C	19 Mar. 99	30 June. 02	1,196.00	912.00	192,002,109.84	225,000,000.00	135.80
4	Gewane - Mille	C	19 Mar. 99	30 Oct.,02	1,218.00	912.00	216,353,945.75	350,305,821.01	146.03
5	Logia - Semera (Contract 1)	C	24 Feb. 95	36 months	1,080.00	720.00	74,643,378.57	88,200,000.00	41.00
6	Tarma Ber - Kombolcha+Check	C	28 Nov. 2000	07 March. 07	2,294.00	900.00	457,812,822.26	529,867,929.24	187.00
7	Woldiya - Alemata (DBST)	C	19 March. 99	14 Nov. 05	2,425.00	1,096.00	150,329,634.00	226,421,169.16	78.30
8	B/Mariam - Wukero (DBST)	C	19 March. 99	1 Sep. 04	1,987.00	1,096.00	203,410,054.65	211,588,154.43	273.26
9	Addis - Gohatsion Section 1 (5.5 - 31.5 km)	C	29 Oct. 99	March. 04	522.00	522.00	58,937,356.28	58,937,356.28	26.00
10	Addis - Gohatsion Section 2 (31.5 - 76.5 km)	C	29 Oct. 99	March. 04	522.00	522.00	58,937,356.28	58,937,356.28	45.00
11	Addis - Gohatsion Section 3 (76.5 - 95.0 km)	C	29 Oct. 99	March. 05	871.00	871.00	58,937,356.28	58,937,356.28	18.50
12	Debre Markos - Merawi	C	19 March. 99	15 Feb. 05	2,156.00	1,096.00	327,070,830.00	412,366,861.18	219.80
13	Merawi – Gondar	SC	10 Oct. 01	31 Dec. 06	1,906.00	1,096.00	395,575,960.57	455,860,533.11	208.00
14	Mekenajo - Nejo (DBST)	SC	20 Aug. 04	3 June. 07	1,018.00	720.00	138,601,822.00	151,414,167.27	61.00

15	Nejo - Medi (DBST)	SC	20 Aug. 04	1 April. 07	896.00	720.00	147,768,639.37	165,089,053.53	70.00
16	Alemgena - Butajira (DBST)	C	25 Oct. 2000		1,247.00	1,472.00	223,651,990.00	237,889,773.88	118.00
17	Butajira - Husana (DBST)	Ongoing	7 July. 2004	30 April. 07	1,028.00	912.00	217,465,178.66	174,076,527.70	100.00
18	Modjo - Awash Arba	C	19 March. 99	31 Dec. 04	2,102.00	1,096.00	227,449,546.00	254,539,566.00	160.30
19	Awash – Hirna	C	01 June. 99		1,887.00	1,096.00	256,542,439.10	297,426,124.01	140.40
20	Hirna – Kulubi	C	13 Mar. 99	25 Sep. 05	2,382.00	1,203.00	188,084,957.20	273,023,657.27	91.00
21	Kulubi - Dengego - Dengeno - Dire Dawa (DBST)	C	1 Jun. 99	51.9 months	1,557.00	913.00	162,178,583.00	204,561,000.00	80.00

Appendix G: Minimum and Required, Highly Recommended Data Set by ROCKS-Road Costs Knowledge System Data

Section 1 through 5 Cost and Work Description	
Reccord	
Country	
Region	
User Classification 1	
User Classification 2	
User Classification 3	
Project or Source Name	
Section	
Report Type	
Report Location	
Cost Date	
Cost Type	**Currancy Code**
Cost Source	**US$ Exchange Rate**
Cost Financing	**Construction Method**
Contract Type	**Economic Cost Factor**
Procurement Method	**Taxes (Y/N)**
Contractor Type	**Contingencies (Y/N)**
Work Description	
Work Type	
Work Predominant Activity	
Section 6- Road Work Cost Per Length	
Base Cost (M$)	**Length (Km)**
Physical Contigencies (M$)	**Duration (months)**
Price Contigencies (M$)	**Total Cost Per Kilometer ($/Km)**
Total Cost (M$)	Total Cost Per Kilometer (2013 $/Km)
Section 7-Road Work Cost Per Area	
	Total Cost Per Carrageway Area ($/m2)
	Total Cost Per Carrageway Area (2011$/m2)
Section 8-Road Work Characteristics	
Pavement Width (m)	Climate Type

Shoulder Width (m) ____	Pavement Type ____
Number of Lanes ____	Surface Class ____
Resurfacing Layer Thickness (mm) ____	Surface Type ____
Depth of Milling (mm) ____	Surface Material ____
Surface Thickness (mm) ____	Base Material ____
Base Thickness (mm) ____	Patching or Crack Sealing Unit Cost ($/m2) ____
Subbase Thickness (mm) ____	Patching or Crack Seal. Quantity Unit Cost (m2/Km-year) ____
Structural Number ____	Spot Regravelling or Regravell. Unit Cost ($/m3) ____
Terrian Type	Spot Regravelling Quantity (m3/Km-year) ____
Section 9-Road Work Cost Details (% of Road Work Cost)	
Mobilization (%) ____	
Demobilization and Site Clearance (%) ____	
Earth Works (%) ____	
Drainage, P Works, Culv.&Min. Bridges (%) ____	Drainage and Protective Works (%)
Major Bridges and Structures(%) ____	Culverts and Minor Bridges (%)
Accesses and Junctions (%) ____	
Pavement Courses (%) ____	Subbase (%)
Shoulder Works (%) ____	Base (%)
Line Marking and Signs (%) ____	Subbase and Base (%)
Environment (%)	Surface (%)
Dayworks (%) ____	
Other (%) ____	
Contingencies (%) ____	
Taxes (%) ____	
Total (%)	
Labor (%) ____	
Equipment (%) ____	
Materials (%) ____	
Total (%)	
Foreign Currancy (%)	Foreign Currancy Code
Section 10-Other Additional Road Agency Costs (% of Road Work Cost)	
Land Acquisition (%) ____	Design (%)
Resettlment (%) ____	Suppervision (%)
Land Acquisition and Resettlment (%)	

Minimum required data set | Highly recommended data set

Appendix H: Elemental Break Down for Testing Regression Models to Estimate Costs of Road Construction Projects.

<table>
<tr><th>Work type</th><th>Work description</th><th>Specific characteristics</th></tr>
<tr><td rowspan="9">Earth work</td><td>Site clearance and/or demolishing</td><td rowspan="8">*Soft material
*Normal excavation
*Intermediate excavation
*Hard excavation</td></tr>
<tr><td>Topsoil striping</td></tr>
<tr><td>*Cut to fill</td></tr>
<tr><td>*Cut to waste</td></tr>
<tr><td>*Borrow to fill</td></tr>
<tr><td>*Imported fill</td></tr>
<tr><td>*Re-soiling</td></tr>
<tr><td>*Temporary earthworks</td></tr>
<tr><td>*Capping layer</td><td></td></tr>
<tr><td rowspan="11">Pavement and surfacing</td><td>Sub-grade stabilization</td><td>*G_{20} compact 95%-100%</td></tr>
<tr><td>Sub-grade preparation</td><td></td></tr>
<tr><td>*Sub-base</td><td>*Gravel type I, II designation $G_{7,}$</td></tr>
<tr><td>*Base course</td><td>*Gravel wearing courses and gravel shoulders</td></tr>
<tr><td rowspan="6">*Surfacing</td><td>Cheap seal prime coat of MC-30</td></tr>
<tr><td>*A/C (40mm-50mm thick layer)</td></tr>
<tr><td>Bituminous and surface courses</td></tr>
<tr><td>Single surface treatment</td></tr>
<tr><td>Double seal surface treatments</td></tr>
<tr><td>*Triple surface treatment</td></tr>
<tr><td>*Upgrade existing carriage way(s)</td><td></td></tr>
<tr><td rowspan="8">Furniture works/Miscellaneous</td><td rowspan="8">Ancillary roadwork</td><td>Gabions</td></tr>
<tr><td>Guardrails on steel psts</td></tr>
<tr><td>Sign faces with paints and background, symbols, lettering and boarders</td></tr>
<tr><td>Preparation of areas for grassing</td></tr>
<tr><td>Reflexive paints</td></tr>
<tr><td>Curb stones</td></tr>
<tr><td>Retaining wall concrete</td></tr>
<tr><td>Side walk concrete blocks with tiles</td></tr>
</table>

* denotes characteristics included during analysis